2017-9

Spatial Thinking in the Engineering Curriculum: an Investigation of the Relationship Between Problem Solving and Spatial Skills Among Engineering Students.

Gavin Duffy
Technological University Dublin, gavin.duffy@tudublin.ie

Spatial thinking in the engineering curriculum: an investigation of the relationship between problem solving and spatial skills among engineering students.

A thesis submitted to Dublin Institute of Technology in fulfilment of the requirements for the degree of Doctor of Philosophy

By

Gavin Duffy, BE (Chem.), M. Eng. Sc., M. Sc. Applied e-Learning

School of Electrical & Electronic Engineering,

Dublin Institute of Technology,

Kevin Street, Dublin 8, Ireland

September 2017

Supervisor: Professor Brian Bowe, College of Engineering & Built Environment, DIT

Abstract

Long considered a primary factor of intelligence, spatial ability has been shown to correlate strongly with success in engineering education, yet is rarely included as a learning outcome in engineering programmes. A clearer understanding of how and why spatial ability impacts on performance in science, technology, engineering and mathematics (STEM) subjects would allow educators to determine if spatial skills development merits greater priority in STEM curricula. The aim of this study is to help inform that debate by shedding new light on the role of spatial thinking in STEM learning and allow teaching practice and curriculum design to be informed by evidence based research.

A cross cutting theme in STEM education – problem solving – is examined with respect to its relationship with spatial ability. Several research questions were addressed that related to the role and relevance of spatial ability to first year engineering education and, more specifically, the manner in which spatial ability is manifest in the representation and solution of word story problems in mathematics. Working with samples of engineering students in Ireland and the United States, data were collected in the form of responses to spatial ability tests and problem solving exercises in the areas of mathematics and electric circuits. Following a pilot study to select and refine a set of mathematical story problems a mixed methods design was followed in which data were first analysed using quantitative methods to highlight phenomena that were then explored using an interpretive approach.

With regard to engineering education in general, it was found that spatial ability cannot be assumed to improve as students progress through an engineering programme and that spatial ability is highly relevant to assessments that require reasoning about concepts, novel scenarios and problems but can remain hidden in overall course grades possibly due to an emphasis on assessing rote learning. With regard to problem solving, spatial ability was found to have a significant relationship with the problem representation step but not with the solution step. Those with high levels of spatial ability were more able to apply linguistic and schematic knowledge to the problem representation phase which led to higher success rates in translating word statements to mathematical form.

Declaration

I certify that this thesis which I now submit for examination for the award of PhD, is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work.

This thesis was prepared according to the regulations for postgraduate study by research of the Dublin Institute of Technology (DIT) and has not been submitted in whole or in part for another award in any other third level institution.

The work reported on in this thesis conforms to the principles and requirements of the DIT's guidelines for ethics in research.

DIT has permission to keep, lend or copy this thesis in whole or in part, on condition that any such use of the material of the thesis be duly acknowledged.

Signature: ______________________ Date: 9-Feb-2018

Gavin Duffy, BE (Chem.), M. Eng. Sc., M. Sc. Applied e-Learning

Acknowledgements

To Brian for his always insightful comments, probing questions and support; to Sheryl for her kindness and generosity in all matters, research and otherwise; to Steve for his kindness and assistance in collecting data and moving couches; to colleagues in CREATE for collaborations and discussions; to the students who participated in this study; and to colleagues in my school for the various ways I was supported during the project.

I am indebted to my own family who were uprooted to spend a year in Ohio while I worked on this project; it was a lot to ask and I very much appreciate the sacrifice they made for me. As always, my parents were extremely supportive in so many ways and I am very grateful to them.

I gratefully acknowledge the support of the National Science Foundation for this work through grant number DRL-1535356.

I gratefully acknowledge the fee support I have received from DIT while registered as a PhD student.

Table of contents

Table of illustrations and figures

Chapter 1 Introduction

Problem solving is embedded in the engineering identity – engineers are problem solvers. Embodied in accreditation criteria and curriculum design, problem solving is a goal of engineering education that is manifest in learning and teaching activities throughout an engineering course.

Engineers are expected to engage in solving many different types of problems that vary in complexity and context and may even lie outside their discipline comfort zone. Problems are by definition scenarios that are new to the engineer, he or she has not solved them before, and this demands in the moment thinking first and the application of memorised practical and technical knowledge second. From making simple decisions so a pipe fitter can deal with the unexpected presence of a wall that was not on the drawing to developing solutions for global warming, engineers are expected to confront and address problems, not shy away from them.

While the terms problem and problem solving have many definitions and have been studied in many contexts, a problem can be considered to have three important characteristics: some initial state of conditions, a goal or final state and, most importantly, an apparent obstacle that prevents the path from initial to final states from being revealed and must be discovered (Mayer, 1992). The challenge to the engineer is to arrange and examine the initial conditions, form some mental representation of the problem in order to find a solution path and solve the problem.

Problem solving requires thinking that is directed towards the problem and, therefore, is a measure of intelligence which, according to some current theories, primarily consists of spatial, verbal and quantitative abilities (Kyllonen, 1996; Sternberg, 2004a). While all abilities are relevant, the focus in this study is on spatial ability and its role in problem solving as this has received little attention in the literature and is poorly understood. It appears that spatial ability is relevant to the problem solving process in the context of science, technology, engineering and mathematics (STEM) education where the problems are related to the assessment and learning of course material. Studies that have compared spatial ability levels to performance in physics (Kozhevnikov, Motes, & Hegarty, 2007), chemistry (Bodner, 2015) and mathematics (Casey, Nuttall, Pezaris, & Benbow, 1995) tend to highlight tasks that are non-routine and that require either reasoning activity or problem solving as the tasks most significantly correlated with spatial ability.

The main objective of this study is to provide a detailed description of the relationship between spatial ability and problem solving and to do so in the context of engineering education using problems that are relevant to this discipline. Greater knowledge of this relationship would allow a more informed evaluation of the role of spatial ability in engineering education, contribute to efforts to address the development of problem solving skills among engineering students and help instructors to understand the challenges faced by the sizeable minority of engineering students who have weak spatial ability.

1.1 Background and rationale

"I stand at the window of a railway carriage which is travelling uniformly, and drop a stone on the embankment, without throwing it. Then, disregarding the influence of the air resistance, I see the stone descend in a straight line. A pedestrian who observes the misdeed from the footpath notices that the stone falls to earth in a parabolic curve". (Einstein, 1920, p. 8)

Many famous scientists have been categorised as highly visual thinkers (Lohman, 1993). Einstein conducted thought experiments in which he visualized scenarios of relative motion to test and refine his special theory of relativity. His visualizations came first, they preceded and were the basis for the mathematical models. His mental capacity was so renowned that 'to be an Einstein' has become a figure of speech to describe anyone with gifted ability in science and mathematics. Yet Einstein was not regarded as a mathematician (Isaacson, 2007); in the early stage of his career he avoided the mathematical development of his theory, instead seeking assistance from a friend and colleague to conduct this aspect of his work for him while he focused on visually based thought experiments. It appears that Einstein had a very high level of spatial ability that allowed him to solve one of the major scientific challenges of the time for it was from the development of his theory of relativity that the problem of Mercury's perihelion was finally explained.

The importance of spatial ability to achievement in STEM education is well illustrated by the following sets of findings. First, students who enjoy taking STEM courses in secondary/high school have higher levels of spatial ability than those who do not (Shea, Lubinski, & Benbow, 2001); i.e. one is much more likely to enjoy a course in science in mathematics if one has strong spatial ability. Second, spatial ability levels developed by high school age are accurate predictors of (i) selecting a STEM course in higher education and (ii) the level to which the STEM education course is pursued. This finding was established by Wai, Lubinski & Benbow (Wai, Lubinski, & Benbow, 2009) who analysed data collected from approximately 400,000 high school students in the US. Those who decided to pursue an education and career in engineering had significantly higher levels of spatial ability in high school than those who were destined to study humanities and social science (HSS). Among the engineering group, those who continued to PhD level were from the very highest in spatial ability ranking in high school. One cannot 'be an Einstein' without having very strong spatial ability.

While the preference and talent for STEM education of strong spatial ability students seems clearly established, less is known about where these students outperform those with lower levels of spatial ability, some of whom also choose to take STEM courses. Studies that have compared spatial ability levels to performance in physics (Kozhevnikov et al., 2007), chemistry (Bodner, 2015) and mathematics (Casey et al., 1995) tend to highlight tasks that are non-routine and that require either reasoning activity or problem solving as the aspects for which spatial ability matters the most. In the case of physics, for example, performance on a reasoning test of Newtonian mechanics was found to have a significant relationship with spatial ability. In chemistry, assessments that were relevant to spatial ability were mostly tests of non-routine problem solving and questions related to crystal structure. In maths, questions that were less procedural and had high imagery ratings were answered more successful by those with higher levels of spatial ability. Problem solving is a theme that appears to be common to these studies.

Less clear, however, is exact nature of the relationship between spatial ability and problem solving. While each of these studies appears to have elements of problem solving in common, they differed in subject matter and methodology. A valid and reliable instrument was used by Kozhevnikov et al. (2007) to measure physics knowledge in their study whereas much of the data collected by Bodner and colleagues (Bodner, 2015) were from course tests and the motivation for the study by Casey et al. was to explain the gender gap in mathematical college entrance test scores rather than to examine the role of spatial ability in problem solving. Findings from such studies highlight a relationship between spatial ability and problem solving rather than providing a detailed description of it. Also unclear is how performance on problem solving tasks varies with spatial ability level. For example, how the behaviour of strong visualizers is different to that of weak visualizers when solving non-routine problems. The purpose of this study was to examine in detail the significance of the relationship between spatial ability and problem solving and to explain how approaches to problem solving varied with spatial ability.

Problems can vary from ill to well-structured, as illustrated in a taxonomy provided by (Jonassen, 2010) who placed story problems at the well-structure end and design problems at the ill-structured end. Problems can be similarly classified as convergent or divergent. Many real world engineering problems can be described as divergent in that several valid solutions can be provided. Story problems are typically convergent in that there is only one valid solution. All types, well and ill-structured, convergent and divergent, are relevant to and are found in the engineering curriculum. In order to achieve the objectives of this study a problem type was required that could uncover the relationship between spatial ability and problem solving but minimise the number of potentially confounding variables, primarily prior knowledge but also including affective issues such as motivation and self-efficacy.

Story problems in mathematics have been widely used to study the problem solving process and this literature contains useful ideas for studying the role of spatial ability in problem solving. Story problems are convergent which allows solutions to be scored as correct or incorrect so that statistical comparisons with measures of spatial ability can be made. The mathematical prior knowledge required to solve simple story problems is covered by the early years of post-primary school (by age 15) and one can either assume the playing field is level and that engineering students know the math procedures to answer these questions or, as was done in this case, the math procedures can be easily tested and ruled out as a confounding variable. Story problems can be designed to be completed in 5 minutes or so thereby allowing several to be administered to participants in 30 minutes to one hour. Like divergent problems, they also require thinking that is directed towards the problem and reveal something useful about the problem solving process. Story problems were selected as the most appropriate type of problem to be used in this study to provide a measure of problem solving ability that could be compared with spatial ability.

The main objectives that were set for this project to achieve are as follows:

- Measure the correlation between spatial ability and ability to solve simple word problems in mathematics among samples of first year engineering students,
- Separate problem solving into two phases, problem representation and problem solution and compare performance on each phase with spatial ability and

- Describe the approaches taken by weak and strong visualizers to problem representation and solution and examine the differences in these approaches to explain why spatial ability is related to problem solving.

An aim of this research is that the findings are relevant to engineering educators. Problem solving is embedded in the engineering education accreditation criteria and engineers are stereotyped as good problem solvers. However, the problems used must be deemed by engineers to be relevant to engineering. It is hoped that by using story problems in mathematics, problems that most engineering educators would expect their students to solve without difficulty, the findings will be valued by the engineering education community which will lead to a greater dissemination of the role of spatial ability in engineering education. Another aim is to mitigate the challenged faced by those who enter engineering courses with low levels of spatial ability. This group can represent a sizeable minority of first year students and are retained at lower rates in engineering programmes than those with high levels of spatial ability (Sorby, 2009). A greater understanding of the spatial-problem solving relationship would allow more focused intervention to help these students. Finally, those interested in improving problem solving skills of engineers should also find this study to be relevant to their efforts.

1.2 Context and outline of thesis

The literature review, presented in the next chapter, discusses three strands of research that are connected through this study – theories of intelligence from which spatial ability as a construct emanated, studies examining the role of spatial ability in higher education and studies and theories of mathematical problem solving. The purpose of this chapter is to clarify the research problem and shape the research questions.

Doctoral research projects are very individual undertakings, with each student bringing his or her own attitudes to bear on the project, not only in methodology but also at the more fundamental level of epistemology. This is not so important in scientific studies for which procedures are defined and must be followed. In social and behavioural science studies such as this project, the research topic can often be addressed through several competing epistemologies, ontologies and methodologies, each justified in its own way. In order for the reader to understand why a research problem was addressed through answering the selected research questions and methodology it is helpful to communicate the philosophical position or, as described by Crotty (1998), the theoretical perspective that guided the study. As outlined in Chapter 3, because this is a study of spatial ability one is immediately tied to the epistemology of the cognitive factors theories of intelligence from which spatial ability was born. A positivist approach is evident throughout the study, particularly so in the first half, while interpretation plays a greater role in Chapters 6 and 7 and the rationale for this is outlined in Chapter 3.

Several preliminary investigations were conducted in the first phase of research on this project and focused on three issues: the role of spatial ability in the electrical engineering curriculum, the importance of spatial ability to performance on college entrance tests and the relationship between spatial ability and conceptual understanding of simple direct current (DC) electric circuits. As a factor of intelligence, spatial ability may be relevant in many areas of engineering education and the purpose of the preliminary work was to highlight some of these areas so that one or two could be selected for a more in-depth study. The findings from this work along with the research questions it addressed and the methodology followed are presented in Chapter 4. Both story problems in mathematics and conceptual understanding of DC circuits were found to be related to spatial ability and the former was selected as the basis for the more in-depth study.

Chapters 5, 6 and 7 constitute this in-depth study in which tests of spatial ability, math story problem solving and basic math competencies were administered to two samples of first year engineering students, one at Dublin Institute of Technology and the other at Ohio State University. A detailed description of the samples and tests are provided in Chapter 5 along with findings from a pilot study that was conducted in the development of the set of math problems and questions that were administered to these samples. Statistical results are presented beginning with the properties of the data distributions and the samples and then followed with results of correlations between the measures. The purpose of the work presented in Chapter 5 is to separate problem representation and problem solving and use statistical methods to measure the significance and strength of the relationship between each aspect of problem solving and spatial ability. It is found that spatial ability is significantly related to problem solving among first year engineering students but not to questions that test the core math competencies needed to solve the problems and, therefore, spatial ability is relevant to the problem representation step.

The next step is to attempt to explain why spatial ability and problem representation are related or why strong visualizers or more successful at representing problems than weak visualizers. Taking each problem in turn, participants' solutions were interpreted based on a coding scheme that is based on a knowledge framework for mathematical problem solving (Mayer, 1992) in order to identify key differences in approach to problem solving between weak and strong visualizers. In Chapter 6, this method of interpretation is illustrated in detail for one of the math problems and results are presented for all problems. Findings from Chapter 6 are the key differences between weak and strong visualizers in problem representation for each of the problems.

Consistency in representing problems is examined in Chapter 7 by selecting a key representation code for each problem, adding these up for each participant and comparing this score to the spatial ability score. Key representation codes were also grouped as being either linguistic or schematic in nature and this allowed another comparison with spatial ability. For some of the problems, the representation challenge was mostly linguistic in nature as a schema was self-evident, for others the challenge was mostly schematic and for one problem representation could be avoided by taking a guess and check approach. It was possible to separately measure the significance of the relationship between spatial ability and each of these types of problem and rank them accordingly.

Findings from the entire study, which represent what has been achieved in this project, are listed in Chapter 8 and analysed in light of the issues raised in the literature review. The extent to which they were expected or unexpected findings based on previously published studies is discussed. The extent to which new knowledge has been created of the role of spatial ability in engineering education and the relationship between spatial ability and problem representation and solving is debated. Also evaluated is the research design; having completed the study, the way in which it was conducted is analysed to identify strengths and weaknesses. Finally, a review of both the nature of the findings and the research design raises several new research questions to be answered in the future.

Summary

Spatial ability is regarded by psychologists as one of three primary factors of intelligence and has been shown to play an important role in many aspects of STEM thinking and education. However, it's currency in engineering education is undervalued relative to mathematical and even verbal ability - college entrance policies in the US typically assess these latter two abilities while mathematics is formally developed throughout the engineering curriculum and verbal skills can often be addressed through writing courses.

Based on studies conducted in physics, chemistry and mathematics, it appears that spatial ability plays an important role in tasks of problem solving and non-routine assessments of subject knowledge. In solving problems, one must first think about the problem in order to develop a mental representation of it from which a solution path can then emerge. The main purpose of this project is to examine the relationship between spatial ability and problem representation in the context of engineering students solving mathematical story problems.

Chapter 2 Literature Review

“Once relegated to lower order processing and concrete thought, spatial abilities are now understood as important for higher order thinking in science and mathematics, for the ability to generate and appreciate metaphor in language, and for creativity in many domains”. (Lohman, 1993, p. 112)

The importance of spatial ability as a factor of intelligence was highlighted by (Wai, Lubinski, & Benbow (2009) in their analysis of data collected from Project Talent, a study that administered a battery of psychometric tests to 400,000 high school participants conducted over a one week period in the US in the 1960s (Anon., n.d.). Included in the design was a follow up questionnaire distributed 11 years later to determine which, if any, higher education degree awards had been obtained by the participants. Wai et al. (2009) gathered verbal, math and spatial scores from the sample and grouped the Project Talent sample based on higher education discipline and level of award. They found marked differences in high school ability profiles between those who were destined to pursue higher education in fields such as engineering versus humanities. Compared to the humanities group, a large difference in math ability was found in favour of the engineering group but the biggest difference was revealed by the spatial measure. Together, math and spatial abilities appeared to act as a filter for both career choice and success in higher education in the 1960s.

Project Talent reflects the North American tradition of studying intelligence and its research design illustrates several aspects of the ontology and epistemology held by this tradition. Sometimes labelled as differential psychology, it is based on the view that intelligence can be understood by measuring individual differences across the full range of human abilities. This is done by administering a wide range of psychometric tests to a large number of subjects and then searching for parsimony in these data using factor analytic techniques. Hence the large sample size in Project Talent and the week long period during which tests were administered. The tests themselves were psychometric in nature reflecting the view that intelligence can be measured objectively and, typically, through paper and pencil tests in multiple choice format. North American studies such as Project Talent reflect the cognitive factors or psychometric view of intelligence from which spatial ability, as a factor of intelligence, emerged. To study spatial ability is to assume intelligence consists of objectively measurable abilities, spatial being one of them.

While other ontologies, epistemologies and traditions exist, the most relevant to this study is that of cognitive factors. A discussion of this tradition is the starting point for this chapter in order to highlight what has emerged from it in terms of an understanding of spatial ability and also what has yet to be resolved and remains contentious. Although Project Talent provides some certainty with regard to relationship between spatial ability and success in STEM education, it did not extend to an analysis of where in the STEM curriculum this relationship is most prominently manifest. Studies that have examined this relationship in specific STEM content areas are discussed in the second section of this literature review. Correlations have been found between measures of spatial ability and several measures of performance in physics, chemistry and mathematics. A common thread from these studies is the ability to solve problems that are non-routine, a theme picked up in the final section of the chapter which reviews the literature on problem solving research, particularly in the context of simple mathematical story problems of the type used in this research project.

2.1 Spatial ability and intelligence/Factors of intelligence

Developed by Charles Spearman, "*the founding father of the study of abilities*" (Dennis & Tapsfield, 1996, p. xi), factor analysis solved the problem of distilling a large data set into a small number of summary findings. It allowed Spearman and those who followed in this tradition to find parsimony in data that consisted of responses from a large number of subjects to a large number of ability tests. He found that performance on any test of mental ability could be explained by two components or factors – general and specific (Spearman, 1904, 1927). The general factor accounted for the significant amount of inter-correlation between all ability tests for any one individual while the specific factor explained the variation that was unique to each test. An individual's performance on one test can predict a significant portion of his/her performance on another test with some people having more of this factor than others, i.e. it varies between individuals but for any one person it is a constant factor underlying performance across a range of tests, hence the term general or simply 'g'. Variation at the individual ability level was explained by the specific ability factor – an individual can have different levels of each specific factor (e.g. verbal, spatial and quantitative) and these also vary between individuals. Therefore, in Spearman's theory, an individual's performance on a test of spatial ability depends on their 'g' factor and on their specific spatial ability factor.

Working from the same epistemology, Thurstone (1938) proposed in his Theory of Primary Mental Abilities that intelligence is better described as consisting of seven principal factors – verbal comprehension, verbal fluency, number, perceptual speed, inductive reasoning, spatial visualization and memory. Since Thurstone, spatial ability has played a prominent role in this tradition of psychology and is found to repeatedly surface as a primary factor of ability along with verbal and quantitative abilities whenever theories of this kind are proposed (Kyllonen, 1996).

While the cognitive factors tradition has been effective in identifying key factors it can fail to describe them in psychological terms. For example, Spearman (1904, 1927) was never more than speculative as to what a psychological description of 'g' should be and, therefore, could not explain what caused it to vary, which is why it retained the non-descript label of 'g'. Although he did suggest it is described as a 'mental energy' which some possess more of than others, his achievement was the development of factor analysis and its application to create the two-factor theory of intelligence. While the concept of ‘g’ is still current, descriptions and explanations of both ‘g’ and specific factors are in a continued state of flux which illustrates both the challenge faced in the cognitive factors tradition of explaining what the factors are, as opposed to their existence, and the dynamic, changing nature of the models and theories within this tradition. So it goes for any of the factors of intelligence, including spatial ability – the certainty of their existence is not yet accompanied by a certainty in their description.

Another view, also current in the North American tradition, is that human abilities have dynamic properties, distinguished as either fluid or crystallized (Sternberg, 2004b) and variation in some tests can be better described by including this distinction. Fluid ability is exposed in tests of reasoning or non-routine, in the moment thinking, while crystallized intelligence is knowledge that has accumulated over time. This distinction is supported by Carroll’s Three-Stratum Theory of Cognitive Abilities (Carroll, 1996) which is based on a large number of data sets collected across several decades, i.e. a very large data set, and also incorporates Spearman’s ‘g’ theory.

Information processing theories of intelligence also exist which hold that variations in performance can be explained by differences in functional aspects of the brain such as short term working memory and long term memory . Evidence for such functionality comes from ability tests and from those who have suffered brain damage in some way and have, for example, lost the ability to remember what happened one hour ago but have no difficulty remembering something from several years ago. Kyllonen (1996) found working memory capacity to have higher correlations than any other cognitive factor with a range of cognitive tests and suggested that it is ‘g’, the general factor of intelligence that impacts on every aspect of cognitive performance.

Yet another theory, based on an information processing cognitive model, is that cognitive tasks require interplay between perception, working memory, long term memory and motor processing (Kyllonen, 1996). As illustrated in Figure 2-1, it is because of individual differences in the capacities of each of these processes that we find variation among individuals in performing all cognitive tasks, including spatial ability tests. Working memory has been shown to be particularly limited (Miller, 1956) with the consequence that working memory overload can have a very significant effect on cognitive performance. In the model developed by Baddeley & Hitch (1974), working memory is conceived to consist of three components – the central executive which is like attention, the phonological loop responsible for holding speech information and the visuospatial sketch pad which is responsible for the storage and manipulation of visual and spatial information in working memory (Eysenck, 2001). When distracted by having to rapidly repeat the word 'see-saw' i.e. placing a demand on the phonological loop) the performance of chess players was unaffected but when asked to press keyboard keys in a clockwise fashion (i.e. placing a demand on the visuospatial sketchpad) their performance was significantly worse (Robbins et al., 1996). Such findings are taken as supporting the distinction of visuospatial and non-visuospatial processing in working memory.

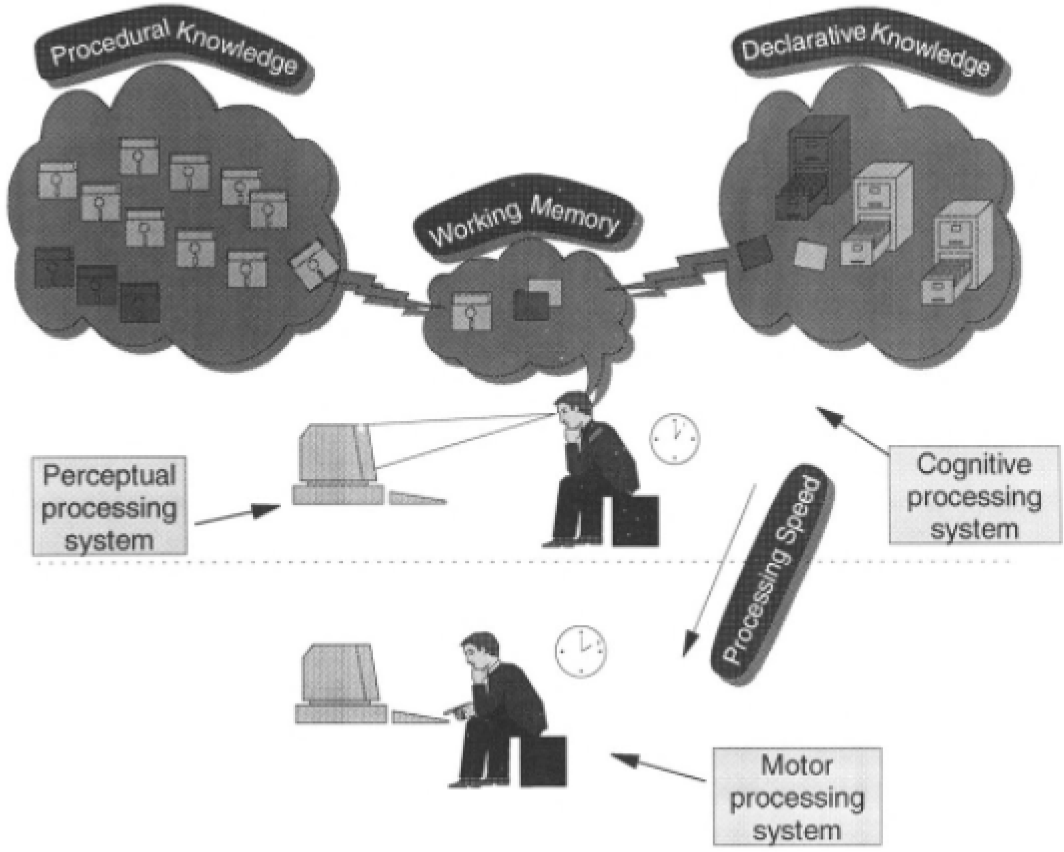


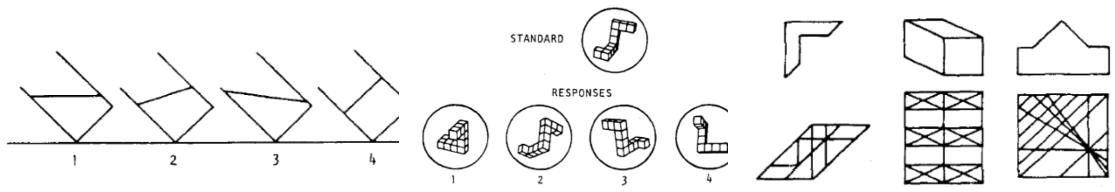
Figure 2-1. Diagram of a general information processing model highlighting aspects that reveal individual differences (taken from Kyllonen, 1996, p. 54).

Spatial ability is an outcome of the cognitive factors theory of intelligence and first came to prominence in Thurstone's Theory of Primary Mental Abilities (1938). Studies of spatial ability are naturally placed within this tradition whose epistemology is manifest in studies such as Project Talent in which the ability tests used are administered in a prescribed format to large sample sizes. While ontological differences abound, most notably in the information processing view of intelligence, spatial thinking appears to remain as a significant characteristic of intelligence.

Defining spatial ability

Identified by the individual differences research as a unique and separately measurable cognitive factor, spatial ability has become a research topic independent of the broader pursuit to understand intelligence. What sets it apart somewhat from the other factors of intelligence is the large and repeatable gender gap in spatial ability in favour of males (e.g. Lippa, Collaer, & Peters, 2010). The quest to understand why males repeatedly outperform females on tests of spatial ability has attracted many to the topic leading to a larger body of research than might not have happened otherwise. To explain the gender gap, some have focused on nature, looking back at human evolution during the hunter-gatherer phase to check for a relationship between spatial ability and throwing accuracy on the one hand (male, hunter) and object location memory (female, gatherer) on the other (Silverman, Choi, & Peters, 2007). Others have focused on nurture to explore the influence of types of childhood play on the development of spatial ability and test, for example, if playing with Lego blocks as a child has an influence on spatial ability as an adult (e.g. Doyle, Voyer, & Cherney, 2012). While the gender gap in spatial ability is not a focus of this project, the literature describing these efforts at solving the gender gap puzzle is a useful source to learn about the ontology of spatial ability.

Through a meta-analysis of a decade of literature on spatial ability research from 1974 to 1985, Linn & Petersen (1985) proposed it consists of three categories: spatial perception, mental rotation and spatial visualization. Spatial perception relates to the way we perceive the world relative to our physical presence within it. For example, the ability to correctly discern the angle of a surface relative to the horizontal or vertical requires spatial perception and is measured by the Water Levels Test Figure 2-2. Mental rotation is the ability to rotate two or three dimensional objects quickly and correctly by imagining their rotation without the aid of real models or sketching the rotation (Figure 2-2). Arguably lacking some clarity in the description provided in their paper, the third category, spatial visualization, appears similar to mental rotation in that it requires visualization to carry out a spatial processing activity but the activity is not a rotational one. Examples of this activity relate to pattern recognition (Figure 2-2), matching two representations of the same object where, for example, one is an unfolded two dimensional version of a three dimensional object. Linn & Petersen (1985) found the gender gap in favour of males was revealed in tests of the first two categories - spatial perception and mental rotation – but not in the third, spatial visualization.



(a) Water levels test/spatial perception (b)Mental rotation test/mental rotation (c) Embedded figures tes/spatial visualisation

Figure 2-2. Sample spatial tests for each spatial ability factor. (Linn & Petersen, 1985)

Where Linn & Petersen discerned three factors, others decided on either different factors, a different number of factors or both. McGee (1979) suggested two factors – spatial visualization and spatial orientation. Lohman (1979) argues there are three – the previous two plus spatial relation. A 2x2 classification of spatial skills is offered by Uttal et al. (2013) in which the object being considered is either static or dynamic and whether the spatial task requires the participant to consider information intrinsic or extrinsic to the object, as illustrated in Figure 2-3. As part of his very large data analysis in developing his theory of intelligence, Carroll specifically addressed the nature of spatial ability and found it best described by five factors. All these different sets of factors are outlined in Table 2-1. One factor that consistently emerges is mental rotation – the ability to mentally rotate well-structured images and to do so quickly. An argument exists for including dynamic spatial factors in addition to the above static factors that are assessed by tests that contain well-structured, static images (Seery, Buckley, & Delahunty, 2015). There is broad agreement that spatial ability consists of several factors but little agreement as to what these are.

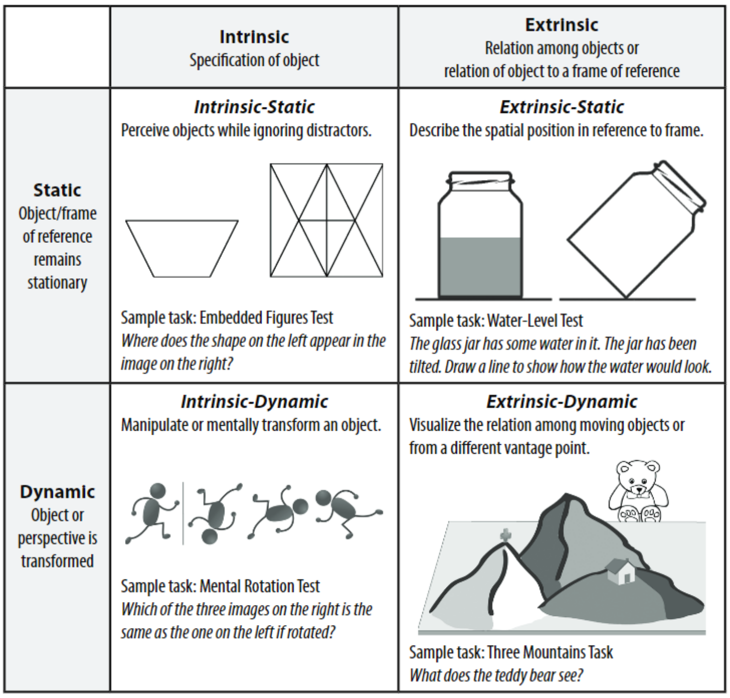


Figure 2-3. A 2 x 2 classification of spatial skills tests (taken from Davis, 2015, p. 16)

In addition to lists of factors, descriptions of spatial ability are also offered by those brave or determined enough to do so. Linn & Petersen suggested that "*spatial ability generally refers to skill in representing, transforming, generating, and recalling symbolic non-linguistic information"* (1985, p. 1482). According to Lohman: "*Spatial ability may be defined as the ability to generate, retain, retrieve, and transform well-structured visual images."* (Lohman, 1993, p. 3). More agreement here, it appears, but these definitions are arguably circumspect and tentative in that someone viewing a spatial test for the first time who is observant enough to spot the well-structured images and the need to mentally transform them could write a similar description. They reflect agreement that spatial ability is defined as mentally processing visual images that do not include tasks related to text or numbers. As responses to the question, 'What is spatial ability?', they are really just a first response and, as Linn & Petersen themselves note, fail to address the problem that "*considerable dispute surrounds the identification of specific spatial abilities and the characterization of the processes used to solve spatial items"* (Linn & Petersen, 1985, p. 1482). And according to Uttal et al. (2013, p.

353),“*unfortunately, the definition of spatial ability is a matter of contention, and a comprehensive account of the underlying processes is not currently available*.” Spatial ability is tested by tasks that require the generation, retention, retrieval and transformation of visual images but the cognitive processes employed while performing these tasks are poorly understood.

Author	Factor name	Description
(McGee MG, 1979)	Spatial Visualization	“the ability to mentally rotate, manipulate, and twist two- and three-dimensional stimulus objects
	Spatial Orientation	the comprehension of the arrangement of elements within a visual stimulus pattern, the aptitude to remain unconfused by the changing orientations in which a spatial configuration may be presented and an ability to determine spatial orientation with respect to one’s body
(Linn & Petersen, 1985)	Spatial Visualization	somewhat tentatively defined as being required for spatial ability tasks that have several steps and can be solved with several strategies
	Mental rotation	the ability to mentally rotate 2 or 3 dimensional images with speed being a key aspect of this factor (to distinguish between analytic and holistic rotation)
	Spatial Perception	“determine spatial relationships with respect to the orientation of their own bodies, in spite of distracting information”
(Carroll 1993 pp362-3, as cited in Lohman, 1993)	Visualization	“Ability in manipulating visual patterns, as indicated by level of difficulty and complexity in visual stimulus material that can be handled successfully, without regard to the speed of task solution”
	Speeded Rotation	“Speed in manipulating relatively simple visual patterns by whatever means (mental rotation, transformation, or otherwise)”
	Closure Speed	“Speed in apprehending and identifying a visual pattern, without knowing in advance what the pattern is, when the pattern is disguised or obscured in some way”
	Closure Flexibility	“Speed in finding, apprehending, and identifying a visual pattern, knowing in advance what is to be apprehended, when the pattern is disguised or obscured in some way”
	Perceptual Speed	“Speed in finding a known visual pattern, or in accurately comparing one or more patterns, in a visual field such that the patterns are not disguised or obscured”

Table 2-1. Spatial factors as identified and defined by several authors.

Leaving aside attempts to define spatial ability and its factors, another way of thinking about spatial ability is to consider why there are individual differences in this ability. Lohman (1993) lists four explanations for individual differences in spatial test performance. The first is based on the finding that time to perform a mental rotation is directly proportional to the size of the angle the object is rotated by which led to the conclusion that mental rotation is an analog to physical rotation (Shepard & Cooper, 1982), i.e. just as it takes longer to physically rotate an object through a greater angle, so too it takes longer to mentally rotate as the angle increases. Therefore, individual differences in spatial ability can be explained by individual differences in mental rotation speed. The premise for the second explanation comes from the working memory hypothesis of the information processing theory: spatial tests may push working memory to its limit which leads to variation in performance among individuals. Third, it is suggested that those with low spatial ability have difficulty in generating and retaining well-structured images whereas high spatial subjects do not. The fourth and final explanation also relates to strategy but is manifest in the way we retain and transform well-structured images. For some, the strategy is holistic in which the entire image is retained and compared to a target image whereas others take an analytical approach by focusing on features of the image and following them through the rotation, if it is a test of rotation (Bodner & Guay, 1997; Linn & Petersen, 1985). The latter is assumed to follow from low spatial ability, the former possible with high spatial ability.

Of these four explanations, Lohman favours the working memory hypothesis – if working memory consists of a phonological loop and visual spatial sketchpad and if spatial ability is governed to a large extent by the latter component, then variation in spatial ability is explained by differences in working memory capacity. A working memory measure can be added to cognitive factor intelligence models. This idea was tested by Kyllonen (1996) who found the best model of his set of psychometric data was neither a purely cognitive factors nor working memory model but a combination of both. This led him to suggest that working memory might be the 'g' that Spearman was certain existed but failed to reveal. Or, as Lohman (1993) argues, if spatial ability is determined by working memory capacity then spatial ability tests might be excellent measures of 'g'.

Measuring spatial ability

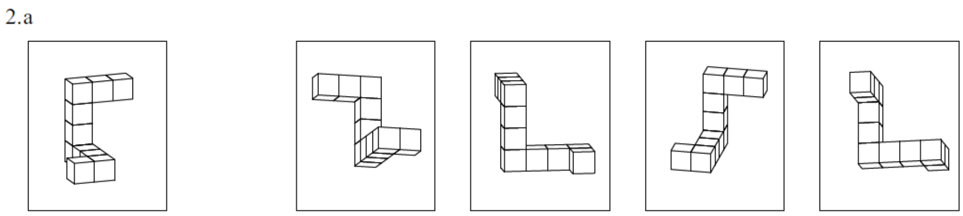
"[Spatial ability] is not a unitary construct. There are, in fact, several spatial abilities, each emphasizing different aspects of the process of image generation, storage, retrieval, and transformation." (Lohman, 1993, p. 3).

Reflecting the contention around the definition of spatial ability there is a plethora of tests of spatial ability. In 1983, Eliot & Smith included 392 different tests in their International Directory of Spatial Tests (Eliot & Smith, 1983) and, no doubt, many more have been produced in the decades since. Following the maxim 'Intelligence is what intelligence tests measure' one can bypass the attempts at defining spatial ability and open a test to get a sense of what skill or ability is assessed by the test. As Lohman (1993) mentioned, spatial ability tests typically consist of well-structured images or figures. Some tests have two-dimensional (2D) objects while others contain three-dimensional (3D) figures. Some contain incomplete patterns, some require the mental transformation of an image while others require mental rotation.

In epistemology, the methods of the cognitive factors or psychometric approach follow the rules of positivism and spatial ability tests must be reliable and valid. Hence, tests contain a sufficiently large number of questions to differentiate guessing from skill. They are typically paper and pencil tests containing a series of multiple choice questions and are completed individually and alone. Large sample sizes are preferred and results are reported for statistical properties of the sample, rather than each individual. Test time is often limited so that both speed and accuracy are required to gain a high score in the test. One motivation for limiting the test time is that visualization can occur as a gestalt or analytical process. Short times discriminate against the latter. There are many tests including pattern recognition, mental rotation, spatial visualisation and mechanical reasoning. Some reveal large gender differences while others do not and it is from observations such as this that arguments for different spatial ability factors are made. Described below are the spatial tests that were used in this study and although they represent a tiny fraction of the 392 or so tests available, they do measure the predominant factors in spatial ability: speeded mental rotation, spatial visualization and spatial perception. These tests fall under the 'intrinsic-dynamic' quadrant in the classification by Uttal et al. (2013) shown in Figure 2-3 which, they argue, are relevant for assessing spatial visualization skills that are needed by scientists.

Spatial visualization/Mental rotation/Speeded Rotation

Mental rotation tests are regarded as the most difficult in terms of visualisation ability as they are difficult to solve through logical reasoning (Peters, Chisholm, & Laeng, 1995). Speeded mental rotation has been widely measured by the Mental Rotations Test (MRT, Peters, Laeng, et al., 1995). The MRT, shown in Figure 2-4, is a redrawn version of the Vandenberg & Kuse (1978) test. It consists of two sets of 12 questions with 3 minutes typically allowed per set. Each question has the same format: a target image is presented on the left with four stimulus images on the right. Two of these are rotated versions of the target and two are not. Subjects are asked to mark the two that are rotated images. Peters et al. (1995) recommend that a point be awarded only if both rotated images are correctly identified so the maximum score per set is 12. It is claimed that the MRT distinguishes between analytical and holistic strategies as the very short time limit only allows those using the holistic strategy to complete it on time (Peters, Chisholm, et al., 1995). Peters, Chisholm, & Laeng (1995) administered the MRT to samples of engineering students and measured scores in the range 13.1 to 14.9 (out of 24) for males and 9.4 to 10.7 for females. When administered again after the passage of one semester both sexes improved their scores by about 3 points. A small number questions on the MRT contain distractors that have a different shape to the target image and can be correctly answered without having to perform a mental rotation. This was examined by Geiser, Lehmann, & Eid (2006) who labelled the 17 % of their sample that answered only these questions correctly as 'non-rotators'.



Instructions: Identify which two of the four images on the right are rotated versions of the target image on the left.

Figure 2-4. Sample question from the MRT-A (Peters, Laeng, et al., 1995).

The PSVT:R (Guay, 1976) consists of 30 multiple choice questions designed to measure 3-D mental rotation ability. The test is timed with 20 minutes normally allowed so the participant's speed and accuracy are assessed. Each question shows a target figure in original and rotated positions. A second 3D shape is then presented in its original position and participants are asked to select from five possible rotated versions the one that matches the same rotation as made by the target. One of the two practice questions from the test is shown in Figure 2-5. This question involves the rotation of the object by 90° around the vertical axis. The participant must apply the same rotation to the second figure and select from one of the five options below a match of the rotated figure. There are 30 questions on the test with rotations varying around one, two and three axes. Reliability measures for the PSVT:R are reported by (Yoon, 2011) with Cronbach's α = .81 measured using data collected from a sample of 180 education major undergraduate students enrolled in mathematics courses. Average scores on this test have been measured to be 21.32 for females and 24.62 for males among a sample of US first year engineering students (Sorby, Casey, Veurink, & Dulaney, 2013).

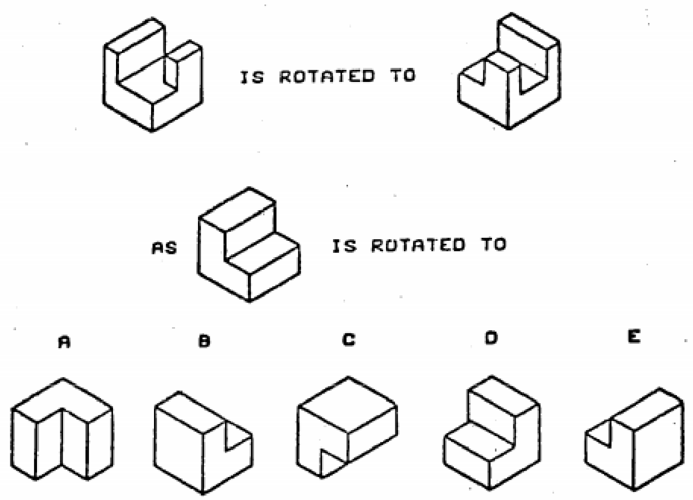


Figure 2-5. Sample question from the PSVT:R (Guay, 1976)

While the MRT is used more frequently than the PSVT:R (Voyer, 2011), the latter is popular in engineering education research (Maeda & Yoon, 2011; Sorby et al., 2013; Sorby & Veurink, 2010b). The PSVT:R has been used as a college entrance test to classify engineering students as either 'weak' or 'strong visualizers' based on threshold score of 60 % (Sorby & Baartmans, 1996), a commonly used percentage to indicate failure on a course in the US. Based on this approach a weak visualizer is defined as 18 or lower on the PSVT:R and strong visualizers as one scoring 19 or higher on the test. There is some debate as to this definition of a strong visualizer with Veurink & Sorby (2011) proposing that those with scores of 19 to 21 (60 to 70 %) be labelled as a 'marginally passing group' rather than strong visualizers. In this study, a decision was made to use the dichotomous arrangement of weak and strong visualizers with the cut-off point as 18 on the PSVT:R. This ensured that all participants or cases were included in statistical comparisons between two samples such as the t-test. Correlation values were also calculated that are independent of any cut off point for weak and strong visualization ability.

In response to the identification of some figural errors in the PSVT:R, Yoon (2011) created a redrawn version of the test. The format, number of questions and number of optional answers are unchanged from the original version as is the recommended time limit of 20 minutes. An example question is provided in Figure 2-6. An average score of 22.87 was measured for a sample of first year engineering students in the US (Maeda, Yoon, Kim-Kang, & Imbrie, 2013).

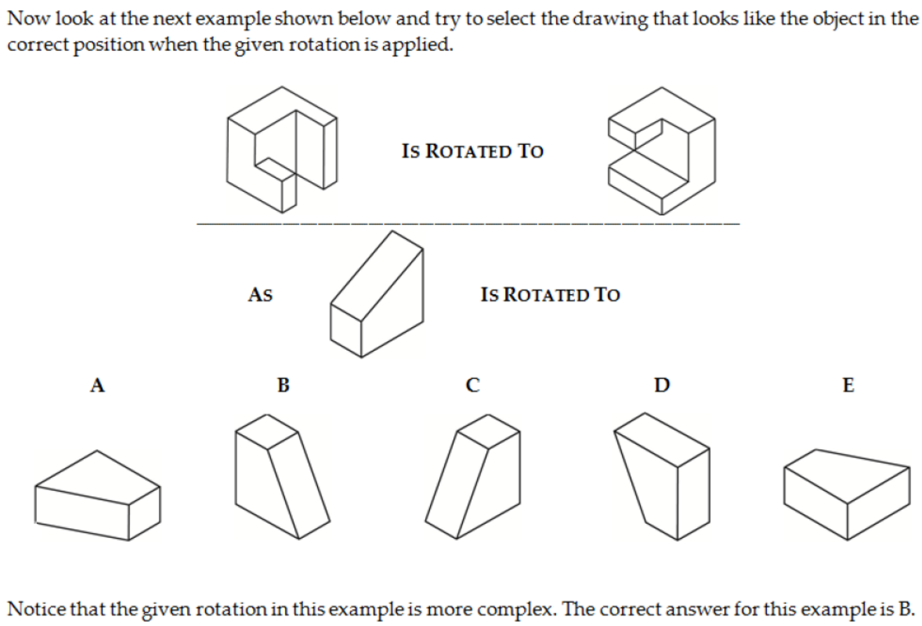


Figure 2-6. Sample question from the Revised PSVT:R (Yoon, 2011)

The Mental Cutting Test, MCT (CEEB, 1939) measures aspects of spatial visualization and spatial relations (Sorby, Leopold, & Gorska, 1999) and is more challenging than the PSVT:R so that MCT data collected from a sample of engineering students will have less skew than PSVT:R data from the same sample. It is not a rotations test but assesses the ability to mentally visualise a hidden face of an object and it also has been used in engineering education research (Maeda et al., 2013; Sorby & Baartmans, 2000). The MCT consists of 25 multiple choice questions with a time limit of 20 minutes normally applied (a sample question is provided in Figure 2-7). Participants are asked to choose one correct answer from five options to match the shape of a cut face of a three dimensional (3D) object. Data collected from engineering and architecture students at a university in Poland yielded an average score of 17.73 for males and 13.03 for females.

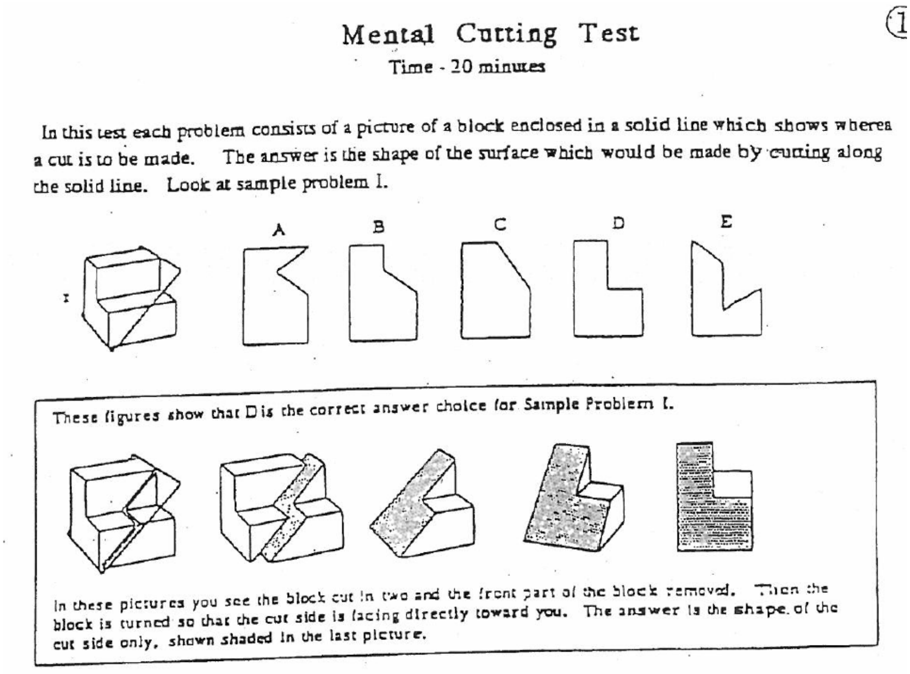


Figure 2-7. Sample question from the MCT (CEEB, 1939)

2.2 Spatial ability in the STEM curriculum

In their analysis of the Project TALENT data set, Wai et al. (2009) created a spatial composite score from four of the tests taken by these students (with weightings in brackets) – three-dimensional spatial visualisation (3.0), two-dimensional visualisation (1.0), mechanical reasoning (1.5) and abstract reasoning (2.0). A sample question from each test can be seen in Figure 2-8. They then grouped the sample by field of higher education pursued (if any) and by highest level of award received (bachelor, master or doctorate) For each group, they then computed a verbal, math and spatial ability score using the data collected while the participants were in high school, i.e. prior to them pursuing higher education.

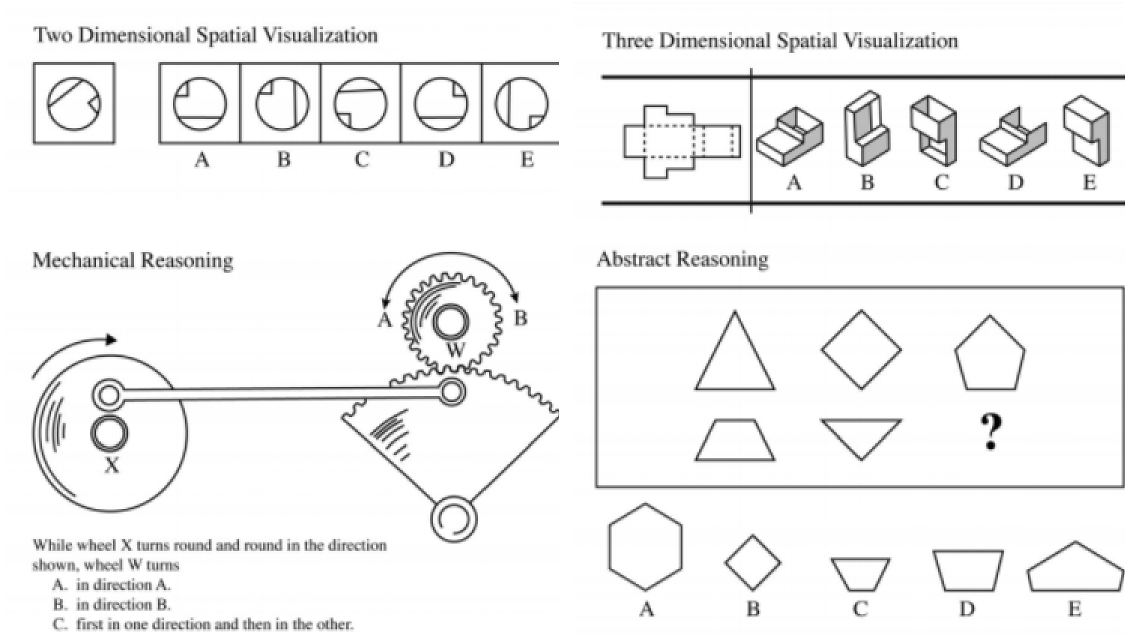


Figure 2-8. Spatial composite used by (Wai et al., 2009).

A graphical presentation of these results, created by Wai et al. (2009), illustrates a distinct difference between the ability profiles of the engineering and humanities groups (Figure 2-9). Verbally, both groups are similar. Mathematically, the engineering group has a much higher score than humanities. With regard to spatial ability, however, not only is the biggest difference revealed but the strength of spatial ability relative to the other two abilities is such that a ‘V’ shaped profile is created for the humanities group while, for the engineers, it is a straight line or ‘I’ shaped profile. The cognitive profile, as measured by the psychometric tests of Project Talent, is markedly different between those who pursued humanities and related social science fields and those who were destined to study engineering and other STEM disciplines. Another point of interest is the positive correlation between ability scores in high school and highest level of award achieved in higher education. Ninety percent of STEM PhD graduates from this sample could be traced back to the top quartile in spatial ability in high school. Having well developed spatial skills as an adolescent indicates a strong likelihood of choosing and being successful in a STEM education and career.

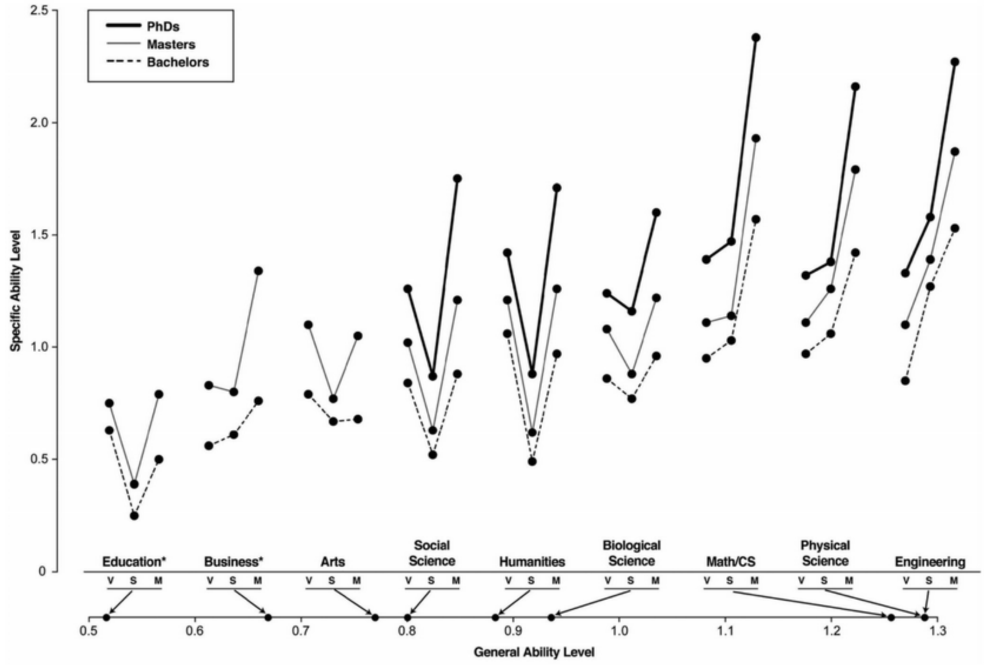


Figure 2-9. Analysis of Project TALENT data to show relative position of spatial scores to verbal and math scores for different disciplines (Figure B1 in Wai et al. (2009))

Given the gender gap in spatial ability it is worth checking the distribution of male and female participants across these categories. As shown in Table 2-2, the numbers of participants in each education level category for STEM and HSS disciplines highlight a prevalence of male students in STEM education in the US of the 1960s and 70s. Therefore, the comparison between STEM and HSS students is a comparison between a predominantly male group and a mixed group. A fairer comparison is, therefore, between the two doctorate groups which are 100 % male for STEM and 82 % male for HSS. Checking Figure 2-9 again with this in mind, the same pattern is revealed – 'V' shaped for HSS and 'I' shaped for STEM doctorates. Although gender undoubtedly has some influence on the results, both male and female students with weak spatial ability are more likely to migrate towards HSS education whereas for those with strong spatial ability, STEM education is the more likely option.

	Bachelor	Master	Doctorate
STEM Male	1137	336	71
STEM Female	6	3	0
HSS Male	1331	390	67
HSS Female	1895	305	15

Table 2-2. Gender distribution across the STEM and HSS groups (computed from data in Appendix A, Wai et al., 2009)

Similar observations were made by Shea, Lubinski, & Benbow (2001) among a sample of academically talented youth who found that those who had developed strong spatial skills by high school will be found in greater numbers among the STEM graduates several years later. They also asked their participants to name the most and least favourite classes they took in high school and reported the findings by gender to allow for potential interaction between gender and spatial ability. For both boys and girls, those who listed STEM classes as their favourite had significantly higher mathematical and spatial abilities and a significantly lower verbal ability than those who selected HSS subjects. Although more boys than girls selected a STEM class as favourite, math and spatial abilities were augmented for both groups. Compared to weak visualizers, those with strong spatial ability migrate towards STEM education and careers in greater numbers, are more likely to enjoy taking these subjects and the education level to which they will study them increases with spatial ability.

Weak visualizers and malleability of spatial skills

Those who select STEM education courses and do not possess strong levels of spatial ability face a greater challenge in completing their studies. The proportion of a freshman engineering class in the US that consists of weak visualizers, i.e. a PSVT:R score of 18 or lower, has been found to be in the range of 10 to 20 % based on data collected between 1996 and 2009 at Michigan Technological University (MTU), (Sorby & Veurink, 2010a). In some years, one in five students were measured to have weak visualisation ability. Women were over represented in this group: the proportion of female weak visualizers varied from 22 and 42 % while for men the range was 8 to 14% (Sorby & Veurink, 2010a). At Dublin Institute of Technology (DIT), the proportion of weak visualizers among the common first year engineering group was found to be 20 % in 2014/15 academic year (Duffy et al., 2015). A sizeable proportion of students join engineering programmes with weak spatial ability and these students arguably face a greater challenge in their studies than their strong spatial ability peers.

At MTU, those identified as weak visualizers at induction stage of first year engineering are offered a course on spatial skills development (Veurink & Sorby, 2011). While this training was initially offered to students it is now a required curriculum component for those who start engineering with weak visualisation skills. The difference in retention rates for those who took the course when it was voluntary compared to those who did not was approximately 13 %. Since the course became a required component improvements in retention have continued to be observed. 'Weak visualizers' are retained at higher rates and get higher grades than those who just passed the spatial test on entry to college (retention: 83 % v 80 %; GPA: 2.83 v 2.64).

STEM curriculum components related to spatial ability

While the studies by Wai et al. (2009) and Shea et al. (2001) highlight the contribution of spatial ability to overall performance in the STEM curriculum, one must look elsewhere for information on which particular curriculum components are most influenced. To be successful in any discipline of higher education requires achievement in a multiplicity of assessments which accumulate to provide an overall grade point average or award classification. The purpose of this section of the literature review is to identify STEM curriculum assessments that strong visualizers outperform those with weak spatial ability. Or, to be more precise, in which subject content areas and in which types of tasks related to these content areas strong visualizers have an advantage. The literature contains studies that have examined the relationship between spatial skills and academic performance in mathematics, chemistry, physics, computer programming, geology and dentistry.

Before reviewing this literature, however, one should be reminded of some basic principles of statistical analysis since these studies typically report statistical correlations. For a correlation value between scores on two measurements to be meaningful, the data collected from these measurements must comply with the following assumptions. First, the measurement instruments must pass some tests of reliability and validity so one can be reassured that the instrument measures what it is claimed to measure and can consistently do so. Second, the data collected should be normally distributed if the subsequent statistical analyses require the data to have this property. When these criteria are not met there is much greater risk of finding a correlation by chance, i.e. a Type I error, or failing to find a correlation where there is an effect, i.e. a Type II error . Probably for pragmatic reasons, many studies that examine the relationship between spatial ability and academic performance use an established psychometric test to measure spatial ability since they are readily available but fail to apply the same standards to the measure of subject knowledge by using course grades. Findings from such studies are, no doubt, of value to the context in which they were measured but they come with a high risk of either Type I or II errors and may create misunderstanding with regard to the relationship between spatial ability and academic performance. When reviewing studies that report statistical findings it is important to remember the principles underlying the statistical methods used.

In a study of 'academically gifted' 1st year STEM students, Miller & Halpern (2013) collected data using a range of spatial tests and compared them with course grades in physics, calculus, chemistry, engineering systems, biology and statistics. A significant correlation was found for the spatial – physics relationship only. While the spatial tests used in this study satisfied a reliability criterion (a Cronbach's alpha is provided), no such information is provided for the course grade measurements. One can only conclude that strong visualizers have an advantage in studying physics but not the other subjects in this context only Although there may be a significant relationship between spatial ability and physics knowledge and skills in general one cannot draw that conclusion from this study. If assessments of the other courses only rewarded rote learning then a Type II error may have occurred: there may indeed be a relationship between spatial ability and aspects of this subject material but it remains hidden by the way the subject material was assessed in this context.

Peters, Chisholm, & Laeng (1995, p. 69) concluded that "*differences in spatial ability as measured by the MRT do not have an impact on course performance*" based on the following findings: male scores on the MRT were 1 standard deviation (SD) higher than female scores while in calculus, female scores were higher than male scores by 0.3 SD and in other subjects, males were slightly better than females but overall, there was no pattern to the differences. They found sex differences in favour of males in spatial test scores on the one hand and no sex differences in academic performance on the other. Again, the instruments they used to measure academic ability, grades on courses and GPA, have unknown reliability and validity and to make such claims in any general sense is misguided. The authors themselves are cautious in interpreting the findings: "*While we do not wish to overstate the significance of the present set of data, they do provide useful caution for those who over interpret claims of any strong link between sex, spatial ability and skills such as are required in mathematics, engineering and similar fields*" (Peters, Chisholm, et al., 1995, pp. 72–73).

Spatial v mathematical abilities

Participants in Project TALENT that had the highest maths scores on the SAT-M also had the highest spatial scores. In Figure 2-9 above, maths scores are offset above spatial scores; as spatial ability increases so does mathematical ability. Over 50 years ago, Smith (1964) argued that spatial skills mediated thinking in mathematics. If so, scores in both should co-vary and this was one of several research questions addressed by Friedman (1992) in a meta-analysis of 136 studies reporting mathematical-spatial correlations. Friedman works from the view that tests of spatial skills in which the participant is presented with different images of the same object can be completed in two ways: spatial 'orientation' refers to a holistic spatial reasoning in which the whole object is visualised during a mental transformation or rotation while ‘visualisation’ relates to the so called analytical approach in which pieces or parts of the object are visualised and a trial and error approach is taken. In her meta-analysis, Friedman’s classification of spatial tests is based on this view and was adapted from work by Schonberger (1976). The classification consists of four categories: 2-D Spatial Orientation, 3-D Spatial Orientation, 2-D Spatial Visualisation, 3-D Spatial Visualisation. The Scholastic Aptitude Test – Mathematics (SAT-M) and Quantitative (SAT-Q) has been used in the US since 1926 as a college entrance examination and has been found to be both reliable and valid (Camara & Echternacht, 2000).

Friedman (1992) found:

1. Mean values of correlations ranging from .35 for 2-D orientation tests to .47 for 3-D visualisation tests (i.e. lumping all studies together and, hence, all different measures of maths)
2. Ranking the four categories of spatial skill used in this study from highest to lowest correlations with maths, the order is:
    - 3-D visualisation, e.g. Differential aptitude test – spatial relations (DAT-SR, highest correlation)
    - 3-D orientation, e.g. MRT
    - 2-D visualisation, e.g. Minnesota paper form board test (MPFBT)
    - 2-D orientation, e.g. Card Rotations Test (lowest correlation)

    I.e. 3-D > 2-D and Visualisation > Orientation. However, differences between them are not large and statistically significant only when highest and lowest (3D V, 2D O) are compared;
3. Reasoning maths tasks (e.g. SAT-Q) have higher correlations with spatial tests than computational maths and more so for older (>14 years old) samples and more so for males;

4. Age did not influence the magnitude of the correlations to any significant extent;
5. Verbal-mathematical correlations are either equal to or higher than spatial-mathematical correlations;
6. Reasoning maths tasks also have higher correlations with verbal ability than computational maths;
7. Visualisation skills have higher correlations with problem solving maths tasks than orientation skills;
8. SAT-Q had remarkably high correlations with spatial when separated from the pool of reasoning maths tasks (r = .67 for 3-D Orientation v SAT-Q);
9. Correlations between spatial tasks and geometry maths are heavily influenced by questions that relate to geometry proofs. When scores related to proofs are omitted the correlations increase significantly. With proofs, geometry correlations are lower than maths in general. Without proofs, geometry correlations are similar to other maths correlations;
10. Contradictions can surface: correlation between 3D orientation v SAT-Q is higher than 3D visualisation v SAT-Q;
11. Females have higher SAT-Q correlations than males. Females tend to have higher verbal-maths than space-maths correlations.

From this meta-analysis she concluded: “correlational studies provide little evidence that spatial skills underlies abstract mathematical thought [and] educators are not likely to be successful in improving performance on mathematical tasks as they are taught and tested today simply by improving spatial skills” (Friedman, 1992, p. 36). Her contradiction of Smith’s assertion that spatial ability mediates math skills is based on the magnitude of the correlations she found in the literature – when lumped together the correlations between mathematics and spatial range from .30 for 2D orientation to .45 for 3D visualization (Friedman, 1995). The reason Friedman appears dismissive of such correlations, which are sizeable for research on human subjects, is that she found higher correlations between verbal and math skills than between spatial and math. Rather than dismiss the role of spatial ability, Friedman could have concluded that both verbal and spatial abilities are important in mathematics. Perhaps such a conclusion was avoided as it shows all the major abilities are correlated and brings one back to Spearman's 'g'. Friedman could also point to the variation in spatial-math correlations based on the different spatial and math measures. For example, the largest correlation was between 3D Orientation (a holistic spatial test) and SAT-Q (a math reasoning task) with $r = .67$, $p < .05$. In addition, correlations between spatial and geometry math increased when proof questions were removed indicating a more nuanced relationship that she appears to realise. Therefore, one can conclude from Friedman's (1992) analysis that spatial-math correlations can be quite high and depend on the nature of each task.

Casey, Nuttall, & Pezaris (2001) measured the spatial ability of 8th grade students in the US using four spatial tests (Vandenberg MRT, DAT-MR, Water Levels Test and a composite of all three) and compared these data with scores on two subsets of the Third International Mathematics and Science Study (TIMSS). TIMSS is a standardised measure of mathematics that has been shown to be reliable (Casey et al., 2001). One subset, TIMSS-male, consisted of 15 questions that males tend to get higher scores on and the other, TIMSS-female, contained 15 questions that females get higher scores on. Questions on TIMSS-male had higher imagery ratings than those on TIMSS-female and the latter were more procedural in nature than TIMSS-male. Imagery ratings for each question were based on the usefulness of using of "*(a) referring to or (b) manipulating pictures in their* [participants] *minds when solving the item*" as judged by five 'experts' in mathematics independently of the authors (Casey et al., 2001, p. 38)There were significant correlations between all four spatial tests and both maths subsets. However, the correlations were much higher with the male maths subset (.44 to .55) than for the female maths subset (.17 to .29). Math self-confidence data were also collected but were found to have a less significant relationship with the two subsets than the spatial – math relationship.

Miller & Halpern (2013) found a significant relationship between the MRT and SAT-M (r(75) = .43). Likewise, Casey, Nuttall, Pezaris, & Benbow (1995) also found correlation coefficients ranging from 0.25 to 0.54 between the MRT (20 item version, Vandenberg & Kuse, 1978) and SAT-M. They collected data from four different samples of students from different academic achievement levels and different age groups: (1) mathematically talented students in 7th through to 9th grade; (2) high ability;(3) low ability (based on SAT-V) high school students who are college bound and in the senior year; and (4) undergraduate university students. A consistent correlation coefficient of 0.35 to 0.38 was observed for the four female samples but there was more variation in the correlations between the MRT and SAT-M for the male students. For males in samples 1 and 4, the correlations were insignificant whereas for samples 2 and 3 they were significant. The authors did search for but could not find reasons to explain the lack of consistency across the male samples.

Study	Math Test	Sample description	r_{male} (n)	r_{female} (n)
Casey et al. (1995)	SAT-M	Mathematically talented 7th to 9th grade	-.03 (102)	.35** (84)
	SAT-M	STEM & HSS undergrads liberal arts	.13 (79)	.35** (195)
	SAT-M	High SAT-V senior high school	.54** (45)	.38** (50)
	SAT-M	Low SAT-V senior high school	.35** (101)	.25** (104)
Miller & Halpern (2013)	SAT-M	STEM undergrads liberal arts	.43** (77, both genders)	
Ganley & Vasilyeva (2011)	Class grades	8th grade	.48* (47)	.11 (67)
	MCAS[a]		.50* (47)	.23 (67)
Friedman (1992)	SAT-Q	Meta-analysis	.67* (large, both genders)	

[a] Massachusetts Comprehensive Assessment System
* significant at p < .05
** significant at p < .01

Table 2-3. Correlations between math ability and spatial ability measured by the MRT for a selection of studies.

Spatial v chemistry

The role of spatial ability in chemistry education has been explored by Bodner and colleagues (Bodner, 2015; Bodner & Guay, 1997) whose studies typically report comparisons between tests of spatial ability and grades in chemistry courses. Chemistry topics whose assessments were found to have a significant relationship with spatial ability included crystal structure in general chemistry, optical activity in organic chemistry and anytime a question required "true problem solving skills" (Bodner & Guay, 1997, p. 5). Pribyl & Bodner (1987) divided their samples into three groups based on spatial ability – low, medium and high. The total sample size was n = 422. In 17 of the 22 exam assessments spread across the four chemistry courses, high spatial ability students achieved significantly higher scores than low spatial ability students. The high spatial ability students had an advantage when mentally manipulating 2-D representations of molecules and when problem solving skills were required. When rote memory or simple algorithms were required there was no difference in the groups. High spatial ability students were much more likely to sketch the structure of a molecule when answering questions. When low spatial ability students did attempt to sketch a structure they often made errors.

Carter, LaRussa, & Bodner (1987) administered the 20 item PSVT:R (Bodner & Guay, 1997) to students on two general chemistry courses, one for agriculture and health science (n = 850), and another science and engineering majors (n = 1648). Each course contained a series of exams – three for the former group, four for the latter – with exam format being multiple choice questions related to the topics covered in the course. Covariance between the spatial test and the exam scores were determined using the Pearson correlation coefficient. The exam questions were grouped by topic and covariance between scores on each topic with the spatial test were also calculated. Correlation coefficients for spatial v exams ranged from .16 to .25 ($p < .001$).

In the study by Carter, LaRussa, & Bodner (1987), the topics that contributed most to the covariance were stoichiometry and factor-label questions such as unit conversions with $r = .29$ ($p < .001$) for the latter. Below is an example of a factor-label question:

"In Apothecaries' measurement, 1 dram = exactly 60 grains and 1 pound = exactly 96 drams. What is the mass in grams of aspirin in a 15.0 grain aspirin tablet? (a) 1.9 x 102 g Co) 2.6 g (c) 1.2 g (d) 0.25 g (e) 5.7 X 10'6 g". (Carter et al., 1987, p. 9)

About 10 % of the variation in performance on questions such as these is shared with scores on the spatial test. Carter et al. (1987) argue that, to the students on the chemistry course, such questions appear to them as problems to be solved rather than routine exercises.

Spatial v physics

Physics education researchers have at their disposal a number of standardised concept inventories which provide a reliable and valid alternative to measuring physics knowledge and skills than course grades. One example is the Force and Motion Concept Evaluation (FMCE, Thornton & Sokoloff, 1998) which is designed to assess the conceptual, qualitative understanding of Newtonian mechanics and has been widely used in physics education research. Kozhevnikov and colleagues have collected data from several samples of students with the FMCE and a number of different spatial tests (Kozhevnikov et al., 2007; Kozhevnikov & Thornton, 2006). Findings from studies that use the same instruments for spatial ability and physics understanding can be compared with each other despite coming from different contexts and provide an insight into the role of spatial ability in STEM education that is more generalizable than comparisons with course grades can offer.

Several studies have shown there to be a significant relationship between spatial ability and the ability to understand and apply the concepts associated with Newtonian mechanics. Using the Paper Folding Test (Ekstrom, French, Harman, & Dermen, 1976) as the spatial ability instrument, Kozhevnikov & Thornton (2006) measured the correlation with the FMCE to be $r(74) = .32, p < .01$. Mac Raighne et al. (2015) also found a significant relationship between the PSVT:R and the FMCE with a correlation of $r(171) = .30, p < .01$. Such moderately sized correlations indicate that about 10% of the variation in tests of conceptual understanding of

Newton's laws of motion, including the ability to interpret graphs of displacement, velocity and acceleration versus time (e.g. FMCE), are shared with spatial ability. Significant correlations were observed in samples of students who were enrolled in a psychology degree and had not studied physics at higher level (Kozhevnikov et al., 2007) and in samples of students who were enrolled in physics at university level but prior to commencing their studies (Kozhevnikov & Thornton, 2006; Mac Raighne et al., 2015). The ability to comprehend and reason about scenarios based on Newtonian mechanics is significantly related to spatial ability.

In order to have a correct conceptual understanding of Newtonian mechanics one must spend time studying this branch of physics and do so in a way that promotes conceptual understanding. Therefore, having high spatial ability does not automatically guarantee a high score on the FMCE as a correlation value might suggest; one must study the subject also. What the findings from these studies do show, however, is that it is very unusual to have a high FMCE score without also having a high spatial ability score. Indeed, the upper left quadrant in a scatter plot of spatial ability versus FMCE was found to be empty indicating that those weak in spatial ability fail to develop a good understanding of Newtonian mechanics (Mac Raighne et al., 2015).

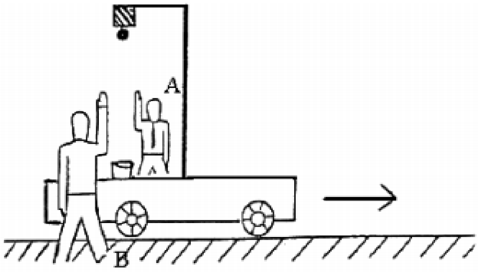
In the Kozhevnikov and Thornton study (2006), spatial and physics data were collected at the beginning and end of a general physics course (n = 76) in which an effort was made to make the lectures interactive and address misconceptions. Of the five problem sets in the FMCE, significant correlations with spatial ability were found with three of them – force graphs, acceleration graphs and collisions – at the beginning of the semester but with only one of the them – acceleration graphs - at the end of the semester. In Question 19 on the FMCE, one is asked to select from eight graphs (or none of the graphs) the one that best matches this description of a car's movement: "*The car moves toward the right, speeding up at a steady rate*." Post instruction, 10 of the low spatial students and none of the high spatial students selected a graph corresponding to a ‘literal’ interpretation of the text, i.e. a line of constant positive slope on a graph with acceleration on the y-axis and time on the x-axis. After instruction, weak visualizers continued to make errors in matching a graph of acceleration over time with a description as shown in Table 2-4.

Spatial Ability	Pre-test Correct	Incorrect[1]	Post-test Correct	Incorrect[1]
Low	9	24	24	10
High	16	16	35	0

[1]. Incorrect defined as choosing the graph that corresponds to a ‘literal’ interpretation of the question.

Table 2-4. Results for FMCE Question 19 reported by (Kozhevnikov & Thornton, 2006)

In another study, Kozhevnikov et al. (2007) used think aloud observations to learn qualitatively how the students approach these problems. In a frame of reference problem shown in Figure 2-10, participants are shown a simple sketch in which a ball is dropped above a cup in a moving car that contains a passenger (one frame of reference) while being watched by a stationary observer (second frame of reference). The think aloud observations revealed how low spatial ability students were much more likely to describe two irreconcilable descriptions of the ball’s trajectory as viewed by each observer. High spatial ability students were more likely to resolve the two frames of reference even though their initial statements revealed similar misconceptions as the low spatial ability students. The sample sizes in this case are too small to provide any statistical significance to the results. The frame of reference observations were based on responses of 11 weak visualizers and 13 strong visualizers drawn from an original sample of 60 undergraduate psychology students who had not studied physics in college or high school. However, it does suggest strong visualizers have an advantage in reasoning through many physics concepts problems.



A small metal ball is being held by a magnet attached to a post on a cart. A cup is on the cart directly below the ball. The cart is moving at a constant speed as shown by the arrow in the figure below. Suppose the ball falls off the magnet while the cart is in motion. Observer A stands on the cart, and observer B stands on the road, directly opposite the post of the cart at the moment of ball releasing.

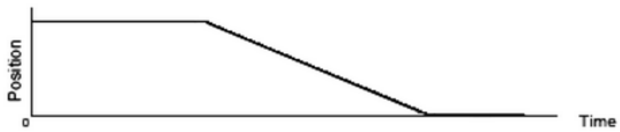
Which of the reports described below corresponds to observer A's view of the falling ball:
(a) The falling ball moves straight down;
(b) The falling ball moves forward;
(c) The falling ball moves backward.

Which of the reports described below corresponds to observer B's view of the falling ball:
(a) The falling ball moves straight down;
(b) The falling ball moves forward;
(c) The falling ball moves backward.

Figure 2-10. The frame of reference problem used by (Kozhevnikov et al., 2007)

Kozhevnikov, et al. (2007) also found that low spatial ability students were significantly more likely than high spatial ability students to interpret graphs as pictures. Interpreting the graph in Figure 2-11 below as a picture would result in response (a), whereas response (d) reflects an ability to more abstractly interpret the graph as data representing the change in position of the object over time. It is assumed the participants had no prior knowledge of such graphs with the assumption justified by their lack of formal physics education in college and high school. Likewise, low spatial ability students were significantly more likely to provide incorrect answers to question 19 to 23 on the FMCE, these questions being related to a graphical representation of a car's motion, even though these students often demonstrated correct conceptual understanding of force and acceleration in other questions.

(a) Here is a graph of an object's motion. Which sentence is a correct interpretation?



(a) The object rolls along a flat surface. Then it rolls forward down a hill, and then finally stops.
(b) The object doesn't move at first. Then it rolls forward down a hill and finally stops.
(c) The object is moving at a constant velocity. Then it slows down and stops.
(d) The object doesn't move at first. Then it moves at constant speed and then finally stops.
(e) The object moves along a flat area, moves backwards down a hill, and then it keeps moving.

Figure 2-11. Graph problem used by (Kozhevnikov, et al., 2007) in think aloud observation

Problem solving is a common theme linking the relationship between spatial ability and the different tasks from maths, physics and chemistry that were found to be significantly related with it. In maths, it was the reasoning tasks assessed by SAT-Q that revealed the highest correlation with spatial ability. In physics, it was questions on the FMCE that required reasoning about physics concepts, scenarios that are very similar to the relative motion thought experiments conducted by Einstein in his development of the special theory of relativity. While in chemistry, it was the non-routine problems that revealed the relationship with spatial ability. Reasoning about non-routine tasks, scenarios and problems appear to draw in some way on the same cognitive features that are assessed by tests of spatial ability. The next and final section of the literature review focuses on the problem solving literature including studies that have examined the relationship between problem solving and spatial ability.

2.3 Problem solving

Those exasperated by the lack of certainty and plethora of meanings associated with spatial ability as a factor of intelligence will find no solace in the literature on problem solving. It too contains many different meanings, theories, interpretations and approaches to learning about problem solving. Problem solving is also a cognitive activity that requires thinking about a task and is, therefore, an aspect of intelligence. It suffers the same definitional, ontological and epistemological challenges as any other aspect of intelligence including the definition of what a problem is and how problem solving should be studied. Complicating matters, perhaps, the overall goal in many problem solving studies is to learn about cognition and intelligence as it can reveal interesting aspects of the way we think. In these cases, problems are used as instruments to measure intelligence rather than problem solving. Intelligence and problem solving become synonymous so to ask 'what is problem solving?' is to ask 'what is intelligence?' Some traditions of research on intelligence can be defined by the ways they have used problem solving activities to learn about intelligence. To avoid returning to a review of theories of intelligence I present two well-known problems from the literature that highlight what has been learnt in general about problem solving. This is followed by a review of the literature on problem solving in mathematics.

The Jug Problem

"You are provided with 3 jugs – an 8 litre jug full of water and a 5 litre and 3 litre jugs, both of which are empty. What is the shortest sequence of pouring water from one jug to another so that you end up with only 4 litres of water in the 8 litre jug. When pouring water into it, each jug must be filled completely as it has no markings to indicate how full it is. One only knows if it is empty or full. For example, you can't tell if the 5 litre jug is half full with 2.5 litres of water."

Figure 2-12. The 'Jug Problem', (image taken from Mustata, n.d.)

A problem can be considered to have three characteristics – some known information that defines the initial state, some goals that define the desired state and, as seen by the problem solver, some impediments to getting the desired state. Problems can be placed somewhere on a continuum between well-structured to ill-structured. Reed (2015) suggests four types of problem – puzzle, insight, classroom and design problems. Puzzle problems are examples of well-structured problems (WSP) and include the Jug Problem shown in Figure 2-12. Problems such as this have been used by Gestalt psychologists to illustrate several aspects of problem solving including *insight*, '*einstellung*' (attitude in German) and *functional fixedness*. All the information to solve the jug problem is provided. One thinks about the problem, possibly not finding any way to solve it until a moment of inspiration arrives and all of sudden the solution path changes from very murky to clear. Luchins (1942, as cited in Mayer, 1992) gave a version of the jug problem to two groups, one who had been trained with similar jug problems and the other that received no training. When presented with a simpler jug problem, the control group (no training) solved it with fewer steps than the experimental group who retained the einstellung they had learned in the practice. Gestalt psychology has shown how important past habits are in problem solving. Failure to solve a problem can be due to entrenched ideas about how to solve it. Letting go of these allows one to move to a different phase of problem solving with new challenges. The importance of attitude was also revealed in work by Bartlett (1958, as cited in Mayer, 1992) and Duncker (1945) who described it as functional fixedness.

The Missionaries and Cannibals (a river crossing problem)

The arrival of the computer presented psychologists with a way to model human thinking as a series of logical steps. For example, Newell & Simon (1972) used think aloud protocols to identify heuristics for how human subjects solved well-structured problems and then showed that a computer program could be created from these heuristics that behaved very similarly to the human subjects. Thinking about a problem is modelled as follows: the *problem solver* has facts, algorithms and heuristics at his/her disposal that can be applied to the *problem space* which consists of the initial (current) state, the goal state, and all possible states in between and result in *operators* which can be physical or mental that result in movement from one problem state to another (Mayer, 1992).

"Three missionaries and three cannibals arrive at a river that they wish to cross. They find a boat but it is only big enough to take two people. If the misionaries on either bank of the river or in the boat are outnumbered at any time by cannibals then the cannibals will overpower and eat the missionaries. Find the simplest schedule of crossings that will permit all the missionaries and cannibals to cross the river safely. When the boat arrives at a bank all the passengers must get off. At least one person must be in the boat for it to cross the river."

Figure 2-13. A river crossing problem with missionaries and cannibals.

For the missionaries and cannibals problem, the problem space includes the problem and the evolving solution. Newell & Simon (1972) proposed that operators are often selected from a limited set of heuristics. One is 'repeat state avoidance' which reflects a desire of the problem solver not to retreat to a previous position and restart the process from there. Another is 'hill climbing' which is to increase the similarity between current and goal states and is revealed by the missionaries and cannibals problem in the step where both a missionary and a cannibal must be moved back to where they started which evokes a feeling of increasing the gap between current and goal state, i.e. going back down the hill you're trying to climb. When guess and check and hill climbing fail to work, a third and more advanced heuristic called means-ends analysis can be applied. In this case, the final goal state is put aside and shorter term goal states are used to resolve obstacles. One goes back down the hill to take a different route to get around the overhanging rock and once this goal state has been achieved then one can see if the final goal state can now be reached. With means-ends analysis one is always trying to climb the hill but detour is an option.

Problem solving in mathematics

College students were asked to translate the following relational statement into a mathematical notation (Clement, Soloway, & Lochhead, 1982): "*There are 6 times as many students as professors at this university*". One third of students incorrectly translated this statement expressing it as 6S = P rather than S = 6P. Simple algebra and word story problems in mathematics can be deceptively difficult. Presented in this section are four studies that illustrate how challenging it can be to correctly interpret problem statements and how subtle changes to a problem statement without altering the mathematical content in any way can lead to noticeable differences in strategy and success rate.

Despite achieving very high grades in mathematics courses, engineering students can struggle to solve relatively simple math problems as Clement (1982) discovered when he administered the following cheesecake and strudel problem (including the instructions) to 150 freshman engineering students:

Write an equation using the variables C and S to represent the following statement:

At Mindy's restaurant, for every four people who ordered cheesecake, there are five people who ordered strudel.

Let C represent the number of cheesecakes and S the number of strudels.

Figure 2-14. Cheesecake and strudel problem (Clement, 1982)

The success rate was 27 % - over two thirds of the class struggled to convert the word statement into the correct equation (5C = 4S). The most common (68 %) incorrect response was 4C = 5S which is obtained by translating one word at a time in the order of appearance without comprehending the entire statement.

To examine this translation process further, (Hudson, 1983, as cited in Mayer, 1992, p. 461) presented two versions of the same problem which differed in the phrasing of the relational statement as shown in Table 2-5 and found very different success rates among two groups of children.

Problem	Pre school	1st Grade
A. There are 5 birds and 3 worms. How many more birds are there than worms?	17 %	64 %
B. There are 5 birds and 3 worms. How many birds won't get a worm?	83 %	100 %

Table 2-5. Success rates for two groups of children on a problem phrased two ways (Hudson, 1983, as cited in Mayer, 1992, p. 461).

Likewise, Hegarty, Mayer & Green (1992) presented the following relational problem in four ways to undergraduate psychology students (Table 2-6). Two versions are phrased to have consistency between the relational statement and the mathematical operation, i.e. 'less' = subtraction and 'more' = addition. Two are inconsistent: 'less' = addition and 'more' = subtraction. The error rate on the consistent problems was 1 % versus 15 % on the inconsistent problems.

Consistent, error rate = 1%	Inconstistent, error rate = 15%
At ARCO gas sells for \$1.13 per gallon. Gas at Chevron is 5 cents less[1] per gallon than gas at ARCO. If you want to buy 5 gallons of gas, how much will you pay at Chevron?	At ARCO gas sells for \$1.13 per gallon. This is 5 cents less[1] per gallon than gas at Chevron. If you want to buy 5 gallons of gas, how much will you pay at Chevron?

[1]. 'Less' is replaced with 'more' to create two more versions of the problem

Table 2-6. Four versions of an arithmetic word problem (Hegarty et al., 1992)

In the final example, Coquin-Viennot & Moreau (2003) presented these two versions of a problem to primary school children (5th grade, age 10 – 11) and observed how the alteration of the problem phrasing was accompanied by a change in solution strategy as shown in Table 2-7.

Problem	Most likely strategy
For a prize-giving the florist prepares for each of the 14 candidates 5 roses and 7 tulips. How many flowers does the florist use in total?	Distributed strategy (14 x 5) + (14 x 7)
For a prize-giving, the florist prepares for each of the 14 candidates a bouquet composed of 5 roses and 7 tulips. How many flowers does the florist use in total?	Factorised strategy 14 x (5 + 7)

Table 2-7. Two versions of an arithmetic word problem (Coquin-Viennot & Moreau, 2003)

These studies illustrate the pitfalls that can occur during problem translation, i.e. translating or converting the word statement(s) to mathematical equation(s). Somewhat like the tourist abroad who mistranslates directions or menu orders, word statements in math problems can be mistranslated causing the answer to be incorrect. Such observations led to a closer examination of how people interpret problem statements in order to learn more about thinking when problem solving. Working with a smaller sample of 15, Clement (1982) used a think aloud protocol to learn more about the thought processes employed when solving the cheesecake and strudel problem. He formed the view that there are two ways in which students made errors: (i) word order matching, in which symbols and numbers are matched based on the order in which they appear in the problem, and (ii) static comparison, in which more thought is given to the statement in an attempt to understand the expression but mistakes are made such as the larger number placed with the larger group or 4C representing 4 cheesecakes rather than 4 times the number of cheesecakes. In contrast, Clement described the successful approach as employing an operation to increase the number of cheesecakes by 5 and the number of strudels by 4. However, some participants performed this correct operation early in their solutions only to later retract it and revert to an incorrect response. To explain why these participants came so close to solving it before turning away, Clement suggested the equality produced by the correct operation is counter intuitive since the restaurant does not possess five times the number of cheesecakes it actually has.

The most common error observed by Hegarty et al. (1992) in the Arco/Chevron problem was a 'reversal error' – selection of the incorrect choice of operation (addition or subtraction) – and this occurred more frequently for inconsistent than consistent problems. This reversal error, it is argued, is explained by direct translation of the statement in English to an equation in mathematics, in the same manner as the word order matching strategy described by Clement (1982). Some participants, however, are immune to changes in phrasing and are able to correctly answer problems that are phrased consistently and inconsistently. These problem solvers don't make reversal errors on inconsistent problems.

Hegarty et al. (1992) concluded from their work that the differentiating factor between success and failure in problem solving was explained by problem representation or mental modelling. Unsuccessful problem solvers did not appear to develop a mental model of the problem; rather, they took a short-cut approach labelled as 'direct translation' by Hegarty et al. (1992) or described as 'compute first and think later' (Stigler et al. 1990, as cited in Hegarty, Mayer, & Monk, 1995, p. 19). In a follow up study of comprehension of word problems using eye-tracking measurement, Hegarty et al. (1995) observed unsuccessful problem solvers spend more time re-reading the problem than successful problem solvers. This extra time was spent looking at the numbers and relational terms in the problem rather than looking at the variable names used in the problem statement. Successful problem solvers were more efficient at processing the portions of the statement that contained relational statements. It is suggested that working memory limitations may lead to the adoption of the direct translation or word order matching approach (Hegarty et al., 1995) as it presumably does not require all the items in the problem statement to be held and rearranged in working memory.

Problem solving in mathematics can be considered to consist of two consecutive phases – problem representation and solution. As explained by Mayer (1992), representation begins with a reading of the problem statement which demands linguistic knowledge to understand the words being used and semantic knowledge to comprehend any 'common knowledge' contained in the problem. Included in the representation stage is the selection of a schema which requires both schematic knowledge and the interpretation of the problem statement to select the appropriate schema. The solution phase can now be entered during which the problem solver chooses the order in which to perform the different mathematical procedures, i.e. a strategy, and for this phase to be successful, sufficient knowledge of the relevant mathematical procedures is required.

"A sleek new blue motorboat travelled downstream in 120 minutes with a current of 8 kilometers per hour. The return upstream trip against the same current took 3 hours. Find the speed of the boat in still water".

Figure 2-15. A distance = rate x time problem (Mayer, 1983, p. 354).

Mayer's model of math problem solving is illustrated with the motorboat problem (1983), shown in Figure 2-15. To solve this problem, one must first represent the problem so that it can be solved using mathematical procedures. There are two steps: representation and solution. Representation has three components - linguistic knowledge of the words in the problem, semantic knowledge (e.g. a river is a body of water flowing between two boundaries) and schema knowledge (distance = rate x time). Solution has two components – procedural knowledge (e.g. arithmetic) and strategic knowledge (ordering the computations).

Errors can be made at any stage in the process, but perhaps less so during problem solution. Procedural knowledge depends on the extent of math education and accomplishment of the problem solver. Math procedures are core competencies that are developed through practice and can be easily recalled assuming the problems are pitched at the right procedural level for the sample, which is typically the case in these studies. However, even when sufficient prior knowledge exists, silly errors are always possible. Strategy is the other element of problem solution in Mayer's model (1992). For relatively simple algebra problems, simple strategies are sufficient, e.g. the hill climbing heuristic, and of all the stages in solving problems of this kind, strategy is arguably the least likely cause of failure. Errors in representation, on the other hand, occur easily as shown in the discussion about translation errors above. Errors in schema development, the other aspect of representation, are also likely to occur.

A series of interesting experiments were conducted by Hinsley, Hayes, & Simon (1977) to tease out the role of schema in solving mathematical story problems. When presented with a large number (76) of such problems from a high school math textbook, participants (n = 14) in one of their experiments were able to categorise the majority of these problems with a high level of agreement. For example, a problem related to 'river current' was placed by many in a different pile to one related to 'distance = rate x time (DRT)'. This showed that participants were quite capable of categorising problems by schema and shared a reasonably common understanding of the schemata. Identification of a schema is somewhat like a moment of insight in problem solving. For example, if a problem is correctly categorised as having a 'river current' schema from memory of solving previous river current problems the task changes from problem to exercise, i.e. it is no longer a problem but a repetition of previous work with different values.

Schema can be identified quickly, even before reading the entire problem. Half of the 6 participants in one experiment correctly categorised the problems after hearing on average one fifth of the problem statement read to them (Hinsley et al., 1977).

Just as problem phrasing can lead to translation errors, so too can it create confusion at the schematic level. In another experiment, they gave a participant problems in which the schema was overt, e.g. DRT, and others where it was disguised, e.g. it appeared as an age problem when it was algebraically DRT. The overt schema problems were solved quickly and in a few steps whereas the disguised schema problems were solved in a 'line by line' approach in which each sentence in the problem statement was converted to an equation without any supervisory level process occurring.

In their fourth experiment, Hinsley et al. (1977) replaced some of the words in the problem statement with made up nonsense words so that although the problem was algebraically sensible, it contained nonsense words. Participants were still successful in solving the problems despite this. As explained by one of the participants in experiment 4: "*What I was trying to do all the time was to try to fit it (the problem) into some normal schema ... As soon as I could pop it into a frame where I could deal with things normally, it was much easier to* *deal with*" (Hinsley et al., 1977, p. 101). Finally, they showed that problems can be phrased to distract participants towards an inappropriate schema. They added directions and angles to a DRT problem were successful in diverting half of 6 participants towards selecting a triangle schema for the problem. Schematic errors can easily occur when reading and interpreting a problem statement.

A commonly used reference for the definition of a problem is Duncker's: "*Whenever one cannot go from the given situation to the desired situation simply by action, then there has to be recourse to thinking*" (Duncker, 1945, p. 1). A key aspect of this statement is that being unable to achieve the desired state by action only depends on the individual in question because what is a problem for one person may not be a problem for another. The problem writer, therefore, cannot know if he has been successful in crafting the problem until it is attempted by the participants as it is always possible that some have seen the problem before (or believe they have seen it before). It is only when the problem is novel that Duncker's or any other definition can be applied. Or, to adopt a phenomenological view, the problem in its appearing must be novel. An essential feature of a problem, one that must exist and cannot be removed for it to remain a problem, is that it is a non-routine activity as experienced by the problem solver.

The errors made by participants in the above studies have at least one common theme – translating the problem statement to a mathematical equation is inappropriately performed by working on one word and/or number at a time rather than thinking of the statement as a whole.

Problem solving in physics

Compared to algebra story problems in mathematics, physics problems require another layer of discipline specific, procedural knowledge, can also require another layer of semantic knowledge if an understanding of physics concepts is required and different schemata are also needed. The problems will appear more complex as a result of these extra components. Consider the problem shown in Figure 2-16 used in a phenomenographic study of physics students' approaches to problem solving (Walsh, 2009). This appears to be a DRT problem but one of the trains is accelerating from a low to high speed which means the DRT schema is not appropriate and a different schema from kinematics is required.

There are two trains, on which the drivers have lost full control of the engines and the two trains may collide at a cross junction. There is a passenger train (Train A) West of the cross junction travelling due East on the track and has a serious engine fault. The driver cannot control the velocity of the train and he informs you of the following:

It is currently travelling at a velocity of 20 ms^{-1} and is 7000 (7 km) metres from the junction. It is accelerating at 0.1 ms^{-2} and the speed of the train will only stop increasing when it reaches it maximum speed of 115 kph (kilometres per hour). The length of the train is 150 m

The Nuclear Waste Train (Train B) is on the other track North of the cross junction and travelling due South. The driver cannot control the velocity of the train and informs you of the following:

It is currently travelling at a constant velocity of 25 ms^{-1} and is 6000 (6 km) metres from the junction. At present, he is unable to speed up or slow down, i.e. he cannot accelerate. The length of the train is 150 m

Will the trains collide? Should they evacuate the town?

Figure 2-16. A physics story problem (taken from Walsh, 2009)

By observing physics students as they solved problems related to Newtonian mechanics, Walsh, Howard, & Bowe (2007) identified several different approaches to solving problems such as the train problem which are listed in increasing order of quality and expertise:

- 'no clear approach' in which the problem is represented as variable(s) followed by what appears to the expert to be a random method of manipulating the variables in the hope of getting an answer and without demonstrating any coherent strategy;

- a 'memory-based approach' in which the problem is represented as an analogous problem encountered previously to identify variable(s) and formula(s) followed by trying to solve the problem as the analogous one was solved;
- an 'unstructured plug and chug' method that represents the problem as a variable or set of variables with related formula(s), followed by a trial and error solution path;
- a 'structured plug and chug' method in which the problem was represented as a formula (or set of) plus variable (or set of), followed by planning a solution based on the variable(s) and formula(s) but often with unforeseen obstacles arising;
- a 'scientific approach' which involved a representation of the problem that included the concepts associated with it, followed by planning and carrying out a solution in a coherent manner.

Those using the scientific approach developed a conceptually coherent representation of the problem that was evident in what guided their solution. Such a representation was missing from the other categories and replaced with other, less appropriate representations with varying levels of success. The next most successful approach, structured plug and chug, worked well but not always. In the transcript example of this category provided in their paper (Walsh et al., 2007), the student demonstrates an understanding of the correct formulas and plans the solution correctly but fails to solve the problem because his representation is not fully coherent. Those in the most basic category, 'no clear approach', fail to identify and commit to any coherent strategy and vacillate to the point that the interviewer begins to makes suggestions, much to the relief, it appears, of the student.

Interestingly, those in the scientific category only demonstrated this approach when "*they were faced with higher-level problems and a strategic approach was necessary*" (Walsh et al., 2007, p. 5) and would otherwise use a plug and chug approach, i.e. when solving a lower-level problem. The level of difficulty of the problem, as it appears to the student, can evoke different categories of approach to problem representation. This may explain the phenomenon, discussed by Uttal & Cohen (2012), that the correlation between spatial ability and academic performance is only found among sample of novices and not among experts. The students in the scientific category were prompted to develop representations when the problems appeared to them to be challenging. When they were not challenging, the development of an appropriate representation was not observed. When the problem appears straightforward and routine, a representation can fail to be observed; when the problem is non-routine, the representation step becomes apparent. The level at which the transition from routine to non-routine occurs is raised as one becomes more expert and, correspondingly, the cognitive demand to develop a representation is lowered to the point where representation ceases to be apparent. If this cognitive demand is facilitated by spatial ability then one should observe spatial ability playing a positive role in non-routine problems which is the territory novices operate within. Hence, spatial ability will matter when studying problem solving among samples of novices but not among experts.

Spatial ability and solving word problems in mathematics

"The strength of spatial ability relative to other abilities, particularly verbal and phonemic fluency abilities, may be more important for predicting how problems are represented and solved rather than whether they can be solved" (Lohman, 1993, p. 4)

A search of academic journal databases for studies that have examined the relationship between spatial ability and problem representation yielded several studies which evaluated the quality of sketches produced during problem solving and the extent to which they assisted in this process. Perhaps the first study of this kind was by Hegarty & Kozhevnikov (1999) who differentiated between a helpful schematic representation and an unhelpful pictorial representation of a math story problem. In their study, 33 boys in 6th grade were administered two spatial tests and presented with a series of math story problems and asked to solve them.

A visual-spatial representation was coded as schematic if a sketch was created that contained schematic content or, if no sketch was produced, a description was provided that included the same content. If the visual-spatial representation was not of this kind it was coded as pictorial. The use of schematic representations was found to have a significant relationship with the math problem score ($r(31) = .48, p < .01$). While a negative correlation was measured between math problem score and the use of a pictorial representation it was not significant ($r(31) = -.34$, N.S.). For example, in the motorboat DRT problem above (Figure 2-15), sketching and colouring a motorboat is a pictorial visual representation of the problem whereas a sketch that contains the DRT schema in some way indicates a schematic representation that will assist in the solution.

Other studies support the existence of a relationship between developing a visual representation that is schematic and success in problem solving. Following a very similar procedure to Hegarty & Kozhevnikov (1999), using the labels, Boonen, van Wesel, Jolles, & van der Schoot (2014) found that those who used an 'accurate visual-schematic representation' were six times more likely to solve a word problems than those who didn't. A study by Edens & Potter (2008) supported the findings of Hegarty & Kozhevnikov (1999); they also found a strong correlation between sketch quality, on a range from pictorial to schematic, and success in problem solving. All these studies were conducted with children in 4^{th} to 6^{th} grade. School children at this level are more likely to be successful at solving word problems if they can visually represent the problem using a sketch that contains schematic aspects of problem representation rather than irrelevant pictorial information contained in the problem.

Spatial ability was also measured in these studies and, for both spatial tests used by Hegarty & Kozhevnikov (1999), found to be significantly correlated with performance on the math problems ($r(31) = .52, p < .01$). The relationship between spatial ability and creating a schematic visual-spatial representation was significant for the Block Design spatial test ($r(31) =$

.36, $p < .05$), but not the Primary Mental Abilities Space subtest, which is a speeded rotation test, $r(31) = -.09$, N.S.). Also, the converse, a pictorial representation, was found to be negatively correlated with both spatial tests but not to a significant extent. In their study, Edens & Potter (2008) found a much stronger correlation between schematic representation and drawing skill ($r(212) = .31$, $p < .01$) than with spatial ability ($r(212) = .13$, $p < .05$). In their study, Boonen et al. (2014) found a positive and significant correlation between spatial ability and problem solving ($r(212) = .59$, $p < .001$) and also between spatial ability and creation of an accurate visual-schematic representation ($r(212) = .31$, $p < .001$). While the findings are somewhat mixed, which may be due to the method of interpreting sketch quality, they do indicate strong visualizers have superior ability to represent a word problem with an appropriate schema, sketch this schema and solve the problem.

It is not clear from these studies whether the representation comes first and is then manifest in the sketch or whether sketching facilitates the development of the representation. It is interesting to compare these findings with Mayer's model which divides representation into three components – linguistic, semantic and schematic – as it appears that some children focus on the linguistic content and sketch that, whereas others can ignore superfluous details and instead concentrate on extracting the mathematical parameters (unknown quantities, known quantities and relationships between them) and finding a schema to relate them.

The view that guided the analysis of the data in this project is that problem representation is a process of translating a word statement into a mathematically coherent form. To effect the translation, the problem solver must first comprehend the statement in linguistic and semantic terms, then identify the connections between the facts and quantities contained in the statement, organise these into relationships while rejecting superfluous information and, finally, select a schema to tie relevant information and relationships together.

Conclusions

Born out of the factor analytic approach to analysing psychometric data in an effort to understand the nature of intelligence, spatial ability is now regarded as a primary factor of intelligence along with verbal and quantitative abilities within this tradition. As an ability, it is manifest in tasks that require the generation, retention, retrieval and transformation of well-structured images. It could be that some are quicker at performing these tasks because they can generate and transform the images in a holistic way whereas others, who have low spatial ability, fail to do this and must operate on parts of the image only. Or, explained from an ontologically different conception of intelligence, the information processing theory, variation in spatial ability is attributed to variation in working memory capacity. Tied to the quest for understanding intelligence, spatial ability remains a tentatively defined concept with many loose ends. This presents a dilemma for the researcher, particularly those new to the field: should one join the quest to precisely define spatial ability or should one pragmatically commit to an informed decision with regard to ontology and epistemology and proceed from there? The latter approach was taken in this study – spatial ability, spatial skills, spatial reasoning and spatial thinking are all synonyms for that aspect of cognition that is measured by intrinsic-dynamic tests such as mental rotation and mental transformation.

In other words, spatial ability is the ability to examine two and three dimensional objects, shown on media such as print or screen as well-structured images, create a mental image of these objects and visualize operations on these images that include rotating them, viewing them from another perspective, unfolding them from three to two dimensions and vice versa, and to do so promptly and without aid from additional resources. This view of the nature of spatial ability informed the research design.

A few points emerge from the literature on overall achievement in STEM education and spatial ability. First, those students who have strong spatial skills tend to enjoy taking STEM subjects and are good at them; STEM courses are arguably attractive to these students and unattractive to their low spatial ability peers. Spatial testing at high school is a good way to predict the future education and career paths of young adolescents – weak visualizers tend to migrate to HSS courses while strong visualizers tend to opt for STEM education. However, engineering educators should expect a freshman engineering class to contain a sizeable minority - up to 20% - of students who have weak spatial skills with a disproportionately large number of them being female. Life in STEM education is arguably more difficult for this group but spatial skills are malleable to some extent and when training is provided, benefits to grades and retention rates in engineering education have been observed. The difficulties these students have appear to be manifest in two areas – activities that require visualization and mental transformation such as visualizing crystal structures in chemistry and in tasks that require reasoning about non-routine problems such as scenarios on the FMCE and story problems in mathematics.

Problem representation is a key initial phase in the process of solving novel, non-routine problems as errors can be easily made in translating problem statements and selecting a schema to guide the subsequent solution phase. While descriptions of inappropriate strategies such as ‘word order matching’ (Clement, 1982) have been developed from observations of problem solving, it is not clear how spatial ability is related to overall strategy. Connections between quality of representation as manifest in a sketch with spatial ability have been examined to show that those who fail to include schematic aspects of the problem in their sketches are unlikely to solve the problem and tend to have low spatial ability (Hegarty & Kozhevnikov, 1999). However, these findings relate to late stage elementary level school children and leave as unresolved the role of spatial ability in problem representation among STEM college students.

A study of the relationship between problem solving and spatial ability among higher education engineering students was therefore merited. The literature revealed that freshman engineering students are challenged to solve 'simple' story problems in mathematics that, for this population, require the application of very basic math competencies implying the challenge is in problem representation and the simplicity is in the subsequent solution. This raised the question as to how problem representation and spatial ability might be related and that this question could be addressed by examining solutions provided by a sample of engineering students to simple story problems in mathematics and comparing approaches to problem solving with level of spatial ability.

Chapter 3 Research Design

"What I mean and what I say is two different things" the BFG [Big Friendly Giant] announced rather grandly. (Dahl, 1984)

We all bring our own likes, dislikes, biases and attitudes to bear on whatever we do, including engineering education research projects, particularly when these are individual undertakings as is the nature of PhD work. The objective of this chapter is twofold. First, it is to communicate to the reader where I stand on the spectra of possibilities in planning education research, these spectra being philosophical in nature as they relate to issues of ontology and epistemology. Second, is to communicate the enactment of these high level assertions by providing details on how I went about the research work. The reader should thus be informed as to what my general attitude to education research is and how I enacted that attitude in this case.

One spectrum of possibility relates to the way we go about creating new knowledge. Academic papers reporting empirical work can vary from phenomenological studies with a small number of *participants* to statistical analysis of psychometric data collected from a large number of *subjects*, with any number of positions in between. The chapter begins with a discussion of two epistemologies – objectivism and constructionism – and illustrates how research work is conducted from each position. The importance of ontology in defining the starting point for a research study is then illustrated using a case study from the research on personal epistemology. In conducting education research, one inevitably enacts an epistemology with regard to research design and an ontology with regard to what should be examined. To avoid behaving like the BFG, researchers must determine where they are positioned on these spectra of possibilities, particularly in education and social research.

A mixed methods design was employed in this study and this was not some random act or convenient choice but reflects a tension between two contrasting positions – the epistemology employed in the cognitive factors tradition from which spatial ability as a construct was developed and the constructionist epistemology that is often more appropriate in education research. The study begins from the cognitive factors position and adopts its post-positivist perspective by using quantitative methods to examine the role of spatial ability in the engineering curriculum and test a hypothesis describing the relationship between spatial ability and math problem solving. A shift in epistemology then occurs as data analysis becomes more interpretive and this is described. The chapter concludes with a description of how the study was conducted in an ethical manner so that every effort was made to protect those who kindly participated in the study.

3.1 Philosophical considerations in social science research

> *'Lost in a remote, rural corner of Kerry, a tourist asks a local farmer for directions to Tralee, where he is staying. The farmer pauses before replying, "If I were you, I wouldn't start from here"'. (Anonymous anecdote).*

The very first step in a research project, before a question is even conceived, is to define a position with respect to what can be known about the phenomenon of interest. How the researcher 'sees' the phenomenon determines not only what questions can be asked about it but how investigations are conducted and what findings can be developed. Those who believe the earth is flat will inevitably write questions that probe its flatness, questions that would never be asked by those who see the earth as a globe. Prior assumptions about the nature of the phenomenon, important as they are in natural science research, are even more critical in human science research where interpretation plays a greater role. Researchers, particularly in the social sciences, should reflect on the assumptions they bring to research and critically examine these assumptions to ensure a justifiable alignment exists between the perspective they bring to their work, the questions they ask and the methods employed (Crotty, 1998). These assumptions should be revealed and made public, rather than remain tacit and undisclosed.

Research methods should be guided by an overarching methodology that is selected based on a theoretical perspective to answer a question which is written from a particular epistemology (Crotty, 1998). As explained by Borrego, Douglas, & Amelink (2009, p. 63), "*the approach used by a given study should be driven by the research questions of that study*". At the same time, the epistemology – what we can possibly learn from this project and how we should go about learning it - determines the way the research question is phrased and how the study is conducted. Linked to this is the theoretical perspective that also guides the nature of the study with perspectives including positivism, interpretivism and critical inquiry (Crotty, 1998). Also included is an ontological position with respect to the nature of the phenomenon – flat earth or globe, for example. These prior assumptions – epistemology, theoretical perspective and ontology – decide the direction of the project, deserve careful consideration and should be clearly defined so readers can fully understand how and why a study was conducted. Like the hapless tourist who asked for directions, one wonders where to start – epistemology, ontology, theoretical perspective or research question.

Epistemology

"'The next thing you know they'll stick me with a behaviourist,' he lamented. 'What do rats and mazes have to do with the mind?'" (Potok, 1995)

Objectivism and constructionism are the dominant epistemologies in the sciences and the humanities and social sciences. Each comes in a separate package which contains a perspective on the nature and permanency of knowledge, a view on how it should be created and a set of procedures to create new knowledge. Epistemology can be deeply personal as revealed by Potok's character, Danny, who had decided to study psychology from reading

Freud but became frustrated by the rigidly scientific approach taken by his teacher. Not all possess Danny's self-awareness and, for many, epistemology can remain unconsidered, subconscious or tacit and lead to inappropriate acts such as the quantitizing of qualitative data (Borrego et al., 2009), somewhat like the BFG.

In the red corner, objectivism embodies the assertion that knowledge exists independent of the human mind, waiting to be discovered by the successful navigator; with time and effort a research question can be answered conclusively and finally. The rigorous, scientific method belongs to this position: through careful, well designed experiments that adjust input variables in a controlled manner and measure the response with accuracy, the question will be answered. Objectivism can lead to 'braggadocio' – scientific facts can and have turned out to be wrong, a point elaborated by Kuhn (1996) in his exposé of serial contradictions of scientific facts that have repeatedly occurred in the history of science; universal truths are replaced by different but also universal truths time after time showing how scientists can fail to adhere to the definition of theory as being tentative in nature and, instead, regard theory as being permanent. Following his review, which was revelatory to him, Kuhn sided with the notion of knowledge being constructed by people and not 'sitting out there' waiting to be discovered, as objectivism would assert. History showed that what was once found 'sitting out there' and accepted as law was often replaced in a revolutionary act with a different understanding and sometimes radically so.

A good example referred to by Kuhn is the dramatic change in the scientific understanding of combustion which occurred during the 1770s when Priestley, a British scientist, and Lavoisier from France made discoveries that could not be reconciled with the theory current at the time. The phlogiston theory, the accepted model in operation for about 100 years at that stage, asserted all combustible materials contained phlogiston which transferred to the air on combustion thereby 'phlogisticating' the air. The theory, however, failed to explain some anomalies such as Magnesium gaining weight during combustion as opposed to losing it as phlogiston departed. Through experimental work, Priestley and Lavoisier cast further doubt on phlogiston by producing an air that allowed the flame to burn brighter thereby highlighting the important role of air in the combustion process. This initiated the demise of phlogiston and its eventual replacement with the view that holds today – the oxygen theory of combustion. These two theories of combustion differ at such a fundamental, ontological level than Kuhn referred to such change as paradigm shift, something one does not readily associate with the objectivist epistemology.

A second message from this story comes from Priestley's resistance to revising his outlook and he departed this world still committed to the phlogiston theory. For whatever reason, he never 'saw' air as consisting of different gases that could be isolated from each other. Instead, he saw the discovery of dephlogisticated air, not oxygen. Lavoisier, on the other hand, did reform his view of the nature of combustion to see it as driven by a gas that was different from air and although he never fully achieved the current scientific understanding he was open to a new reality. Both of these eminent scientists based their conclusions on the same experimental data but came to vastly different conclusions – this is the definition of interpretation in constructionism, not objectivism.

Kuhn provides many such examples in the history of natural science research – a Ptolemaic versus Copernican view of the solar system, mechanics before and after Newton, then the failure by Newtonian physics to explain Mercury's perihelion until Einstein's theory of relativity was developed, to name but a few. As one example of many such stories in the history of science where "*nature has somehow violated the paradigm-induced expectations that govern normal science*" (Kuhn, 1996), the discovery of oxygen shows how science regularly reforms current understanding. While the phlogiston scientists may seem silly by today's standards, they were hard working scientists who did excellent scientific work. How could we and they look at the same fire and see two different things? They viewed combustion from the accepted paradigm of their day and their prior assumptions as to the nature of combustion led them to explore combustion from a very different point of view. Their prior assumptions led to questions that those subscribing to a different theory would not think to ask.

Meanwhile, in the blue corner, constructionism sits in direct contrast to the objectivist position claiming that knowledge is constructed in the human mind through interaction with the environment and with others. Understanding is not sitting out there waiting to be found but is constantly changing and research questions are never resolved. Instead, one simply arrives at a new position knowing more than before. As Crotty explains, "*There is no meaning without a mind. Meaning is not discovered, but constructed*" (Crotty, 1998, pp. 8–9). From this position, one is not obliged to work towards resolving the matter for ever more and this freedom opens the door for other methods and practises that are not found in the objectivist toolbox. This freedom is a very important distinction between objectivism and constructionism.

In science, every phenomenon typically has one accepted model or theory (e.g. oxygen theory of combustion). What constructionism means in this case is that this theory may be reconstructed at a future date, a position also allowed for in the post positivist view. In social science, however, every phenomenon typically has a plethora of competing models or theories each of which is being reconstructed all the time. In this case, constructionism enacted through an interpretive methodology is an arguably more productive approach to take. Objectivism and constructionism provide different views on the research process; neither is right nor wrong but more or less appropriate depending on the context.

Every so often a radically new way of viewing the world is developed. If this is happening in the world of science then how are things in human science research? Here, the phenomena studied are people and their interactions with their own thoughts, with each other, their environment and so on, and findings are inherently more susceptible to change over time than a theory of combustion. If natural science is, at times of radical change, an interpretive process then surely human science research is even more prone to paradigm shift and for such shifts to occur more frequently. Why search for objectivity at all?

Ontology – the role of prior assumptions

Research on epistemological development is one example of human science research that currently has several competing ontologies, i.e. different conceptions of the nature of personal epistemology. Epistemological or intellectual development of college students has been an active area of research since the 1950s (Perry, 1999) with three competing conceptions of what epistemological development is. First, Perry's model (Perry, 1999), and others derived from it (Love & Guthrie, 1999) conceive development to occur along a single path made up of several stages along which students progress from dualism to relativism. A second ontology is to replace the unitary stage model with a multi-dimensional stage model: it asserts that epistemological development occurs on several parallel paths that are independent of each other, each with its own stages (Schommer, 1990). Rather than developing along paths through stages, a third ontology is that students have a large suite of epistemological resources available to call upon depending on context or their expectations of learning. This is known as the framework of resources model (Hammer & Elby, 2002). Each of these three views represents a very different conception of what epistemological development is.

Consider the dilemma faced by those who wish to study engineering students' intellectual development and must choose where they lie with regard to the nature of this phenomenon. Subscribing to the unitary stage model implies one can ask students to think about ill-defined problems unrelated to their engineering programme and then transfer those findings to all contexts, including engineering education. These researchers (e.g. King & Kitchener, 1994; Wise, Lee, Litzinger, Marra, & Palmer, 2004) categorize participants based on their responses to questions such as how can we determine whether genetically modified food is safe or how can we know how the pyramids in Egypt were built. It is then assumed that they will think in similar ways in all areas, including engineering. This lack of context would not occur if the study was based on the framework of resources model as it asserts that epistemology is enacted in context and therefore must be observed in context (e.g. Danielak, Gupta, & Elby, 2014). Those researchers will not ask about the pyramids as they don't believe they can learn about students' thinking in engineering from such questions.

We find that while the same phenomenon – variation in enacted epistemology – is explored by all three positions, each conceives it to have a different form, goes about exploring it in different ways and draws different conclusions as a result. Another example of 'I wouldn't start from here'! How the phenomenon is conceived to exist determines what we can come to know about it. A researcher must carefully make this decision before the project begins as a question written from one ontology makes little sense in another. Hence, the literature should be reviewed not just with regard to what has been discovered but how and under what assumptions.

A mixed methods design

> *"A mixed methods approach is one in which the researcher tends to base knowledge claims on pragmatic grounds (e.g. consequence-oriented, problem-centered, and pluralistic). It employs strategies of inquiry that involve collecting data either simultaneously or sequentially to best understand research problems. The data collection also involves gathering both numeric information (e.g. on instruments) as well as text information (e.g. on interviews) so that the final database represents both quantitative and qualitative information". (Creswell, 2003, pp. 19–20)*

The dilemma I have faced in designing this study is that, on the one hand, I see constructionism as the more appropriate to education research, which rejects a desire to pursue certainty in findings, but, on the other, the research topic itself emanates from the scientific, psychometric tradition; the topic demands a positivist epistemology, at least to begin with. I have dealt with this dilemma by placing a foot in both camps. To begin, the project is guided by a post positivist epistemology through the development of hypotheses, the use of paper and pencil tests, following strict procedures in collecting data and analysing these data using quantitative methods. Having established the existence of a phenomenon by verifying a hypothesis, it then shifts towards a constructionist epistemology by taking an interpretive approach to analysing participants' problem solutions. The dilemma is resolved by beginning the study with a *verification* approach, then changing tack to a '*discover and seek to understand*' mode and finishing by integrating the findings from each.

In summary, this study examines the relationship between spatial ability and problem solving in two ways: (i) measurement of the statistical characteristics of the relationship among samples of engineering students and (ii) examining the qualitative nature of this relationship to understand its features and explain why it exists. The study adopts a cognitive factors ontology of human abilities and examines the spatial-problem solving relationship from two perspectives, post positivism and interpretivism. Several sets of quantitative data are collected for a preliminary analysis of the topic which is then followed by an in-depth study consisting of quantitative and qualitative data which were collected together and are analysed in sequence. As described by Creswell (2003), the study employs transformative and concurrent procedures as both forms of data are collected at the same time, findings from each are integrated, not separated, and a theoretical lens is used (the Mayer framework) as an overarching perspective for interpreting the data.

As for any type of design activity, mixed methods design in human science can and should be evaluated to ensure quality is enhanced. Options for evaluating a mixed methods study include a method orientation, a timing of phases of the investigation orientation and a research process orientation, as outlined by Choudhary & Jesiek (2016). The framework proposed by O'Cathain, Murphy, & Nicholl (2008) is an example of the latter that can be applied to this study. It is presented in Table 3-1 along with an indication of where evidence of meeting each criterion can be found in this report.

Item	Description	Chapter
1	describe the justification for using a mixed methods approach to the research question	3
2	describe the design in terms of the purpose, priority, and sequence of methods	3
3	describe each method in terms of sampling, data collection, and data analysis	3 to 7 incl.
4	describe where integration has occurred, how it has occurred, and who has participated in it	3 to 7 incl
5	describe any limitation of one method associated with the presence of the other method	3, 8
6	describe any insights gained from mixing or integrating methods	8

Table 3-1. Good reporting of a mixed methods study (GRAMMS, O'Cathain et al., 2008)

A mixed methods approach offers the potential to have the best of both worlds. First, results from the quantitative work present an air of certainty as data were collected from large samples of engineering students in two locations and provide reassurances with regard to repeatability. Quantitative results include the verification of a relationship between spatial ability and problem representation which has the benefit of justifying further analysis. Those readers who favour quantitative results, arguably the majority among engineering educators (Borrego et al., 2009), will readily accept these findings as they value this method of research. Hooked by the quantitative results, these readers are likely to continue reading through the qualitative sections. Likewise, those who value constructionist research will readily listen to the findings from the interpretive portion of the project and will realise statistical analysis is mostly of the descriptive kind. Finally, the detailed examination of participants' textual responses that is undertaken during the qualitative analysis helps reduce the risk of Type I and II errors that may have been made in the quantitative phase.

A mixed methods design also bring some threats and risks. A correlation between spatial ability and problem solving does not necessarily mean improvements in spatial ability will cause improvements in problem solving. The existence of a third variable that is common to both spatial ability and problem solving is neither confirmed nor denied by the interpretive phase of the study and causation is left unresolved in this study from a positivist perspective. However, the deeper understanding of the phenomenon created by the interpretive work will provide greater focus should there be a follow up positivist study of causation. Despite these limitations, the benefits offered by a mixed methods design to examining the relationship between spatial ability and problem solving were deemed to be sufficient to justify adopting it as a methodology in this case.

3.2 Procedure/Research process

Research work on this project was divided into several phases beginning with a relatively large phase that was preliminary to the analysis of the relationship between problem solving and spatial ability which was the focus of the subsequent phases. An outline of the work conducted in these phases is provided in Table 3-2**Error! Reference source not found.** along with a summary of the outcomes and predominant methods of analysis used in each phase.

Phase	Work conducted in this phase	Outcomes relevant to next phase	Methods	Chapter
Prelim	The role of spatial thinking in the engineering curriculum was examined by (i) measuring SA in different years of study (ii) comparing SA with math skills (iii) comparing SA with electric circuits knowledge	Justification to pursue the role of spatial ability in electric circuits problem solving and/or math problem solving	Quantitative	4
1	Pilot a set of math problems	Significant relationship between math problem solving and spatial ability	Quantitative	5
2	Administer two math tests, one consisting of a set of simple problems and a second consisting of a set of core competency questions	Significant relationship between spatial ability and the test of problem solving only, therefore spatial ability relevant to problem representation step and not problem solution step.	Quantitative	5
3	Analyse verbal and written responses to each individual math problem based on the Mayer framework – an interpretive process	Greater understanding of how weak and strong visualizers can differ in problem representation	Qualitative & Quantitative	6
4	Compare problem representation across problems, i.e. compare approach in one to the others to search for consistency, common features and differences	Identify trends in problem representation of weak and strong visualizers	Qualitative & Quantitative	7

Table 3-2. Summary of the phases and outcomes of each phase in the research work.

Preliminary work examined the role of spatial ability and spatial skills in the engineering curriculum in three ways: (i) with electrical engineering in DIT as a case study, variation in spatial ability across students in different years of the programme was measured as was its relationship with all modules/courses grades on the programme, (ii) scores on US college entrance tests, include measures of math ability, were compared to spatial ability using a sample of freshman engineering students at OSU and (iii) understanding of electric circuits concepts was compared with spatial ability for samples of first year engineering students at DIT. This work involved the administration of different tests to samples of students or seeking permission to access data from previously administered tests followed by statistical analysis of these quantitative data and all three aspects are described in detail in Chapter 4. The preliminary work phase also provided an opportunity to learn how to conduct research in this area and to reflect what research questions related to problem solving/spatial ability could be addressed in greater detail.

Following the preliminary work, research on the major theme of the study, the spatial-problem representation relationship, began in Ohio State University (OSU) with a pilot study using a small sample of undergraduate teaching assistants in the Department of Engineering Education (EED). This resulted in a set of problems that were administered to a larger sample of first year engineering students at both DIT and OSU. Funding for this phase of the study came from a NSF grant to examine the relationship between spatial ability and problem solving in engineering which included the use of electroencephalography (EEG). Only the OSU students participated in this aspect of the project. The objectives of this research project and those of the NSF funded project overlapped sufficiently to allow data to be collected for both projects concurrently. Hence, data were collected from participants at OSU while they solved problems using both a think aloud and an EEG protocol. The EEG data were analysed by others while this thesis is concerned with the analysis of the written and audio data.

Spatial ability vs. Problem representation, quantitative work

Phases 1 and 2 of the main study were designed to evaluate the statistical significance of the relationship between spatial ability and problem representation. Both the literature review and some preliminary work established that spatial ability correlated with assessments of non-routine reasoning and problem solving tasks in several STEM subjects. Assuming the non-routine component of problem solving is contained in problem representation and not in the solution phase, it was hypothesised that if problem representation could be isolated from problem solution, one would find a correlation between spatial ability and representation but not between spatial ability and problem solution. Isolation was indirectly achieved by creating two math tests, one consisting of a set of problems and the other of a set of core competencies corresponding the solution phase of each of the problems. Logically, problem representation is the difference between spatial-problem solving and spatial-problem solution and this argument is elaborated further in Chapter 5 where both the hypothesis and corresponding null hypothesis are presented.

Instruments

Between the preliminary work and phases 1 and 2, data from six different types of test were collected in this study, as listed in Table 3-3. With regard to spatial ability, four spatial tests - MCT, PSVT:R, Revised PSVT:R, MRT-A - were selected based on relevance to the spatial visualization and rotation factors, ability to discriminate among engineering students and satisfying reliability criteria. The standard protocols for administering all four tests are outlined in Chapter 2 and were followed in this study. In preliminary work, data were collected using an electric circuits concept test called The Determining and Interpreting Resistive Electric Circuit Concepts Test (DIRECT, Engelhardt & Beichner, 2004) and was administered using the protocol recommended by the authors which is described in Chapter 4. Course grade data were also collected as part of preliminary work from both DIT and OSU and, for the latter only, college entrance test data were obtained. Finally, to examine the math problem solving – spatial ability relationship, two math tests were developed for this study. The samples of participants are described in Chapters 4 and 5.

Test	Source	Reliability/Validity
MCT	(CEEB, 1939)	Medium to high reliability, KR-20 = .57 to .64 (Sorby & Baartmans, 2000), KR-20 = .81 (Kelly Jr, 2012)
PSVT:R	(Guay, 1976)	High reliability, internal consistency KR-20 = .74 to .83 (Branoff, 1998)
PSVT:R Revised	(Yoon, 2011)	High reliability, Cronbach's α= .84 (Maeda et al., 2013)
MRT-A	(Peters, Laeng, et al., 1995)	High reliability, split half reliability = .72 to .80 (Geiser et al., 2006)
DIRECT	(Engelhardt & Beichner, 2004)	High reliability, internal consistency KR-20 = .70 (Engelhardt & Beichner, 2004)
MPT	OSU Dept. of Mathematics, unpublished	High reliability, Cronbach's α= .79 (Chapter 4)
SAT Math	www.collegeboard.org	Test is valid (Camara & Echternacht, 2000; Shaw et al., 2016)
ACT Math	www.act.org	Test is reliable (Powers, Li, Suh, & Harris, 2016)
ACT SCIRE	www.act.org	Test is reliable (Powers et al., 2016)
DIT course grades	DIT academic grade book	Not available
OSU GPA	OSU academic grade book	Not available
Math pilot problem	Appendix A (created for this study)	Not available
Math problems	Appendix B (created for this study)	Low reliability, Cronbach's α = .49 (6 items)
Math questions	Appendix B (created for this study)	Medium reliability, Cronbach's α = .61 (6 items)

Table 3-3. List of measurement instruments used in this study.

Quantitative methods

Working in the post positivist epistemology, one must conform to standards of validity and reliability with regard to data collection to strengthen and support generalizability in respect of the findings. Validity and/or reliability data, obtained from the literature where possible or measured as part of this work, are presented in Table 3-3 and discussed below. Also discussed is the nature of the distribution of the data collected from the samples as statistical methods depend, for example, on the data being normally distributed. When conducting statistical analysis of quantitative data, one must adhere to the rules and principles that accompany these methods.

With regard to validity of spatial tests, i.e. the extent to which they measure what they claim to measure, this is a matter of debate that is tied to the ontology of spatial ability and how many factors it contains. As discussed in the literature review, a pragmatic assumption was made that spatial skills that are relevant to STEM education can be measured by tests that fall under the intrinsic-dynamic grouping (Davis, 2015; Uttal et al., 2013) and, based on their widespread use in the literature, tests from this group such as the MRT, PSVT:R and MCT are valid measurements of spatial ability . Tests of reliability assess the extent to which the test agrees with itself and include internal consistency measures such as Cronbach's alpha, Kuder-Richardson 20 and split-half reliability. A test is considered reliable if performance on any one part of the test is predictive of performance on another. If the result of a reliability test is low it indicates that performance is based on several factors and if high, it is more likely the test assesses a single factor. Threshold values above which reliability is accepted are 0.7 for Cronbach's alpha (Field, 2013) and 0.8 for Kuder-Richardson 20 (Sorby & Baartmans, 2000) and all of the spatial tests used in this study have been shown to meet one of these criteria (see Table 3-3). The MCT is arguably the most difficult of these tests with the average scores in samples of engineering students lower for the MCT than other spatial tests (e.g., Farrell et al., 2015). Test difficulty can lead to guessing and, therefore, a reduction in reliability (Sorby & Baartmans, 2000). For example, when administered to younger, middle school students, for whom the test is very challenging, reliability was found to be poor (Hungwe, Sorby, Molzon, Charlesworth, & Wang, 2014). Among STEM students in higher education, the spatial tests used in this study have been measured to be reliable and are accepted as valid tests of spatial ability factors.

As discussed by Engelhardt & Beichner (Engelhardt & Beichner, 2004), DIRECT meets the threshold value of reliability and is also, they argue, a valid instrument for measuring student understanding of direct current resistive electric circuits. The evaluated content validity through feedback from experts who teach the subject on the content of the test and construct validity was evaluated using factor analysis and qualitative analysis of interviews with students as they answered the questions.

Reliability data are provided for the ACT and measurements of validity are provided for both the SAT and ACT which are based on correlations between scores on these tests and college grades (Powers et al., 2016; Shaw et al., 2016). Indeed, these tests are considered as benchmarks in the US education system as they are used to determine eligibility for college entry and are deemed to be accurate predictors of performance in college. Data collected from course grades in DIT and OSU were analysed in the preliminary work but these measurements come without any indication of reliability or validity which undermines the potential to generalise findings. However, they are what matter to students whose final level of award depend on them and issues of validity and reliability are a moot point since there is no alternative measure of academic performance that can be used. The OSU Math Placement Test, also used in the preliminary study, is an unpublished test administered to freshman students during orientation. Since reliability data for this test were not available, a Cronbach's alpha was calculated as part of the work which indicated the test had a high level of reliability.

Finally, both the math problems and questions test used in the main body of the study were found to have a Cronbach's alpha less than .7. They both fail to pass this test, particularly so for the math problems, which implies the 6 questions on the test are measuring several factors rather than one, i.e. performance on one part of the test does not predict performance on another. The implications this has for the results and the generalizability of the findings are discussed in Chapter 8.

Before statistical methods such as correlation and regression were applied to the data, normality of the data distributions were evaluated by plotting a frequency distribution and visually checking it, calculating skewness, kurtosis and their standard errors and conducting tests of normality. The Pearson correlation was used as standard unless the assumption of normality was violated, in which case a Spearman rho correlation was used instead. Simple regression and analysis of variance were calculated using sums of squares which is the standard approach (Field, 2013). All statistical procedures were performed using IBM SPSS version 21.

Some of the key findings from the quantitative work are correlation values that indicate the magnitude or strength of the relationship between two variables and the significance level or probability that these particular data distributions occurred by chance. Although statistical results, correlation magnitudes are subject to interpretation and debate as to what is or isn't large. Cohen (1988) suggested a rule of thumb, shown in Table 3-1, for assigning a qualitative description to a correlation in behavioural science studies which can be referred to when judging correlations reported in this study.

Correlation, r	**Description**	**Variance, r^2**
.10	Small effect	1 %
.30	Medium effect	9 %
.50	Large effect	25 %

Table 3-4. Heuristic for judging correlation magnitudes (Cohen, 1988)

Spatial ability vs. Problem representation - qualitative work

Selecting a quantitative analysis procedure is relatively more straightforward than choosing a qualitative method from the many that are available. A sensible approach is to select a procedure that is a good match with the research question so there is compatibility between the two. Having established the existence of a significant relationship between spatial ability and problem solving among engineering students in phases 1 and 2, new research questions about the spatial-problem representation relationship were now asked in phases 3 and 4. Approach changed from verifying deduced hypotheses to a mode of discovery to learn about the relationship between two cognitive factors, spatial ability and problem representation. This required a qualitative approach that not only matched the new objectives but could also be applied to the data collected and the context in which they were collected. Data consisted primarily of written responses to the six problems collected from all 115 participants in the study and audio recordings of thinking aloud collected from the OSU sample and for three of the six problems. A qualitative method was required that could work with these data and address the research questions.

Dominant qualitative methodologies include ethnography, narrative, phenomenology, grounded theory, action research, discourse analysis, phenomenography and basic qualitative inquiry. Of these, ethnography and narrative research require the researcher to spend extensive amounts of time with the sample, repeatedly observing participants and developing an intimate knowledge of the phenomenon as it exists in their lives, approaches not matched with the research problem in this case. Phenomenology and grounded theory were also not a good match as the assumption of a relationship between spatial ability and problem representation would undermine these methods. Bracketing or epoché in phenomenology requires the setting aside of prior assumptions, such as the role of spatial ability in problem representation, in order to allow freedom for the truth to emerge about what it is like to experience solving word story problems in maths. Likewise, grounded theory emerges from the views of the participants who, unaware of what their factors of intelligence are, cannot describe them. Discourse analysis was also ruled out for the same reasons. Action research is used to improve the researcher’s ability to perform a task and was not relevant. Of the several dominant modes of qualitative inquiry, all apart from phenomenography and basic qualitative inquiry, were redundant as they were not suited to the research question and/or type of data collected.

Phenomenography has been widely used to examine students’ approaches to problem solving (e.g., Irving, 2010; Walsh, 2009; Zoltowski, Oakes, & Cardella, 2012) and is worth considering in this case. While criticised for ignoring the noetic side of experiencing a phenomenon, i.e. the attitude we bring to bear on the experience (Ashworth & Greasley, 2009), it is a method that results in an outcome space which is a hierarchical list of the (assumed to be) limited set of ways of experiencing the phenomenon (Walsh, 2009). Since we learn through variation, we can learn about the phenomenon by observing the various ways in which it is manifest. In this case, there are two phenomena of interest - 'math problem representation' and 'spatial ability testing' - and a phenomenographic method would require the collection of rich descriptions of experiencing these phenomena, i.e. interviewing students about performing each task. Two outcome spaces would then be created from the same set of participants which would allow a relationship between them to be examined using relational phenomenography (Walsh, 2009). Taking this approach to spatial ability could be problematic as interviewing students about taking spatial tests is likely to lead to a verbatim description of their moves (e.g., '*I looked at that corner on the first image and then rotated it and then checked the other part*', and so on) which is unlikely to result in anything other than the already established analytic/holistic dichotomy (e.g. Geiser, Lehmann, Corth, & Eid, 2008). Information on the cognitive processes that are asserted by cognitive psychologists to be occurring in our minds as we undertake these tasks has evaded capture by think aloud protocols. Phenomenography, while a useful method for examining approaches to problem solving, was deemed to be unsuitable in this case given the presence of spatial ability in the research question.

It was decided the most appropriate match to the question and the data was to employ basic qualitative inquiry as implemented by coding the participants solutions to the problems in a quest to understand why there were differences in success rates in the problems, what mistakes were being made, what types of representation were created and how these varied between weak and strong visualizers. Two options to coding were available – the tabula rasa, open out approach and the prior assumptions, focus down approach.

Rather than make prior assumptions about how participants in my sample approach problem solving, the open out approach would start with a blank slate and, through reading the data, generate a set of categories to describe the approaches to problem solving. A benefit of the open out approach is the truthfulness of the codes as they come directly from the data. A drawback is the failure to see something meaningful that is hidden or obscure and which others have seen before when analysing similar data sets. The focus down approach begins with what has been reported in the literature with regard to problem solving approaches. Assuming the model or models provided in the literature are relevant to this context, a set of codes are developed first and then used to analyse the data. While the focus down approach offers the potential to save time which can be made available to do other work, the drawback is assumptions may be inappropriately forced on the data.

I employed both a focus down and an open up approach in the data analysis which I summarise here and illustrate in more detail in Chapter 6. The Mayer framework (1992) outlined in the literature review was used to develop a set of initial codes for analysing problem solutions. Each transcript from each problem (no. of participants = 115, no. of problems = 6) was examined and checked against these codes which were updated as needed to best match the data. These were scored in a binary form – present/absent = 1/0. Analysis first required interpretation of the text and, if available, the audio recording, in order to judge whether a participant employed a particular representation or not. In many cases this was evident and in some cases unclear. I worked alone on this analysis and did not have anyone else available to independently judge the participants' solutions. Hence, no inter-rater reliability data are provided but this was an individual, self-directed study rather than a group collaboration. Where I felt uncertain about a coding decision, I slowed down, considered several options, made a decision and reflected on it. There were times when analysing a transcript near the end of the sample that I realised there was another way of looking at this problem that, in hindsight, was relevant to earlier transcripts. In those cases, I updated the list of codes, returned to the beginning, to the first student, and worked forward through the set once again.

Participants from the OSU sample were asked to think aloud while solving three of the six math problems. Think aloud observation is a concurrent reporting method in which participants are presented with a task and asked to express their thoughts as they come to them concurrent with working on the task. A smart pen was used to both audio record what they verbally expressed and to concurrently record what they wrote on paper. This allowed me to play back a recording later on and listen to what was being concurrently spoken and written down which is very useful in elaborating the thinking that produced the written text. A detailed description of the think aloud procedure is provided in Chapter 5.

An advantage of concurrent over retrospective reporting is a reduced interview and observation time. Participants are not asked to stay on or return for a second session in which they retrospectively report what they were saying as they solved the problems. A drawback to concurrently reporting on one's work is the increased cognitive load the need to verbally express thoughts creates. Articulating thoughts requires working memory which can lead to reduced attention being paid to the problem solving task. Retrospective reporting arguably creates less interference in the participant's performance. However, Van Gog, Paas, Van Merriëmboer, & Witte (2005) found that concurrent reporting yielded more information about the participant's actions, why decisions were made and how problems were being solved than was yielded by retrospective reporting. Hence, think aloud observation is an effective method of collecting data that contains useful information about processes such as problem solving.

Ethical considerations

All stages of this project – research design, data collection, analysis and reporting – were evaluated with ethics in mind in order to ensure participants were protected from harm. Ethical approval for the study was applied for and obtained separately from all institutes in which participants were registered students: DIT, OSU and UNL. A key element of ethical conduct was to fully inform participants about the project and risks and benefits associated with participation before they agreed to participate. By adopting this and other ethical considerations, participants were protected from harm at all stages of this project.

Ethical approval was obtained from the Institutional Review Board (IRB) at UNL which covered data collection at both UNL and OSU. Both institutes adhere to the same rules and principles of ethics so approval at one was accepted by the other. To cover the work undertaken at DIT, an application was submitted to the DIT Research Ethics Committee which granted approval for the project. Think aloud and EEG data were collected in the autumn of 2016 and approval for this work was contained in the original application to the UNL IRB. Ethical approval was also obtained prior to collecting data during the orientation programme for first year engineering at OSU in summer 2016. At all stages, permission to collect data was obtained from the ethics committee or IRB at the respective institute.

Informed consent was obtained from each participant in advance of each data collection. Consent forms outlined the research project, what is expected of the participant, risks and benefits associated with participation (of which there were none) and confidentiality was assured in reporting of data by anonymising participant identity. Also contained in the forms was an explanation that a participant has the right to withdraw from the project at any time for any or no reason. Participants were given time to read the consent forms and to also discuss them if they so wished. Copies of these forms can be found in Appendix E. A UNL IRB requirement was that participants in the US samples had to be aged 19 or older. Participants in the DIT sample were not remunerated for completing the 30 minute math tests as this was administered during a scheduled class time. US participants were required to be available for a longer and more demanding session outside class time which lasted 90 minutes and involved the fitting of an EEG hat. They were remunerated with a gift card of small value for their participation.

All data collected in this study is stored on a password protected computer with access granted only to those who were on the research team for the broader study. This was to ensure participants' privacy was not compromised and to prevent open access to the data which could be used by others with negative consequences that could damage participants and the reputation of engineering education research more generally. Participants will be required for future studies and it is important to establish and maintain trust with students and colleagues by conducting our work in an ethical and professional manner. In the reporting of all data collected in this study, confidentiality has been ensured by using code names for participants rather than real names. Some participants performed well on the math and spatial tests while others achieved low scores and it is important the identity of all is protected. It is not possible to trace data back to identify any individual who participated in the study.

From research design through to reporting, ethical issues were identified and addressed in order to minimise harm to participants in the study. Institutional policies and procedures relating to ethics in research were adopted and followed. Both in spirit and in procedure, the project was conducted in an ethical manner.

Summary

In conducting engineering education research one is faced at a fundamental level with choosing from two contrasting epistemologies, objectivism and constructionism. Neither is right nor wrong but just as there are horses for courses, each has implications for the framing of research questions, the selection of a perspective to guide the study and research methods. Studies of spatial ability are born out of and tied to a cognitive factors ontology of intelligence, a construct created through an objectivist epistemology and studies of spatial ability are naturally pulled in this direction. This study began from this position by examining the role of spatial ability in engineering education in a broad sense and, more narrowly, in problem solving and representation. Having resolved a null hypothesis in the quantitative phase, new questions were posed that probed the qualitative nature of the relationship between spatial ability and problem representation. Theoretical perspective shifted from post positivism to interpretivism before findings from both perspectives were integrated to provide a detailed description of the relationship. Based on such integration, the study can be categorised as a mixed methods design and a method for reporting this type of design was followed.

Several tests of ability were administered in this study including tests of spatial ability and conceptual understanding of electric circuits, grades in college courses, US college entrance tests and, created for this project, tests of math problem solving and core competency. Data collected from these tests were analysed using quantitative methods with attention paid to issues of validity, reliability and other assumptions these methods are based on. Basic qualitative inquiry was selected as the best match from a range of qualitative methods to the research questions asked in the final phases of the project.

The aim of this project is to contribute towards a greater understanding of the role spatial ability plays in STEM education so that curriculum can be evaluated and improved with regard to cognitive development. The main emphasis of the study is to examine the spatial-problem representation relationship and this is addressed in Chapters 5, 6 and 7. Preliminary work examined more broadly the role of spatial ability in engineering education and this work is present in the next chapter. While the gender gap in spatial ability in favour of men is widely reported, the relationship between problem solving and spatial ability was not examined or analysed separately for males and females. Likewise, other demographic properties such as socio-economic status and nationality/ethnicity, which have also been shown to reveal differences in spatial ability (Lippa et al., 2010; Wai et al., 2009), were not included in the research design. It was decided to focus on the spatial ability characteristics of the sample and to examine differences in approach to problem solving between weak and strong visualizers and use these comparisons to learn about the relationship between problem representation and spatial ability.

Chapter 4 Measuring the role of spatial thinking in the engineering curriculum.

Introduction

Research activities in this project were initially focused on learning in a broad way about the role of spatial skills in the engineering curriculum to identify areas that should be pursued in greater detail in the subsequent phases of the project. This preliminary analysis was conducted at two locations - Dublin Institute of Technology (DIT) in Ireland and Ohio State University (OSU) in the USA. The first set of objectives addressed in this chapter relates to the role of spatial thinking across the curriculum. With the Bachelor of Electrical Engineering programme in DIT as the context, it was decided to first measure the extent to which spatial ability changed as students progressed through the programme and to contrast or benchmark this against the Bachelor of Architecture in DIT in the assumption that spatial ability development is a stated outcome of architecture education. In addition, it was also decided to search for aspects of the electrical engineering curriculum at DIT that revealed the strongest correlations with spatial ability. Second and in the OSU context, the objective was to focus on the relationship between spatial and math abilities using data collected from college entrance tests. The last objective was to examine the relationship between spatial skills and the ability to perform reasoning tasks related to simple direct current (DC) electric circuits. It was established in the literature review that strong visualizers can be more adept at addressing problems related to Newtonian physics concepts and this prompted a similar investigation in the context of electric circuits concepts.

The motivation for examining these different aspects of the curriculum and their relationships with spatial ability was to identify and justify the research question addressed in the subsequent phases of the study. By examining a wide range of measures of academic performance in engineering education, from college entrance testing to performance across all years of a programme, and comparing these to measures of spatial ability, it was anticipated that several potential areas for further, more detailed investigation would be identified.

4.1 Spatial thinking in an electrical engineering curriculum

The first objective was to learn about the spatial ability characteristics of the electrical engineering students and to search for curriculum components that revealed a difference in success rates between weak and strong visualizers. It was not initially clear that spatial ability was an important factor in electrical engineering education. To answer this question data from course grades were obtained from the institute's grade database and spatial ability tests were administered to participants from the electrical engineering programme.

Undergraduate electrical engineering programmes in DIT are offered at two levels – engineering technology and bachelor of engineering. Engineering technology is an ab initio three year programme placed on Level 7 of the Irish National Framework of Qualifications (NFQ). The Bachelor of Electrical Engineering is at Level 8 on the NFQ and prior to 2014/15 this was a four year ab initio programme. Since 2014/15 those wishing to pursue the bachelor of electrical engineering must first complete the common first year engineering programme and then progress to complete three years on the electrical engineering programme. Data collection in this project straddled this transition period.

The questions addressed in this section include:

1. How does spatial ability change as students progress through an electrical engineering programme in DIT?
2. Where in the DIT electrical engineering curriculum do differences in grades emerge between weak and strong visualizers? (Weak visualizer defined by a score of 18 or lower on the PSVT:R)

Data collection

During the 2013/14 academic year, students in each of the four years of the Bachelor of Electrical Engineering programme were administered the PSVT:R and MCT tests during normal class time with a one week gap between tests. Module grades for each cohort were collected after the Semester I exams (January 2014) and Semester II exams (June 2014). Average grades for each cohort were calculated taking into account the number of credits for each module. The same spatial ability tests were also administered in the same academic year to students on all five years of the Bachelor of Architecture programme so that a comparison could be made between spatial ability development of architecture and electrical engineering students.

Results & Analysis

The numbers of students in each cohort who took each test along with descriptive statistics are shown in Table 4-1 and Table 4-2 for the two programmes being compared, Bachelor of Electrical Engineering and Bachelor of Architecture. A one way ANOVA with year as factor and spatial test scores as the dependent variables was conducted for both sets of data.

	Electrical Engineering			Architecture		
Year	n	M	SD	n	M	SD
1	22	11.55	4.98	28	11.6	4.2
2	13	10.62	4.50	36	15.1	4.9
3	33	10.58	4.62	29	15.2	4.6
4	14	10.86	6.00	38	15.7	4.4
5				29	17.2	4.2

Table 4-1. Descriptive statistics for MCT data collected from Bachelor of Electrical engineering years 1 to 4 and Bachelor of Architecture years 1 to 5, DIT, 2013-14.

For the electrical engineering samples the result was not significant for either the PSVT:R scores ($F(3,106) = .504$, $p = .681$) or the MCT scores ($F(3,78) = .186$, $p = .906$). This could imply that the number of years spent studying this electrical engineering programme does not appear to have an effect on spatial skills level as measured by both the PSVT:R and the MCT. Since the samples are relatively small and independent it is also possible that a different finding may emerge if one sample was followed over time from first to fourth year with spatial data collected each year. In this instance, all four samples of electrical engineering students appeared to be very similar with regard to spatial ability.

	Electrical Engineering			Architecture		
Year	**n**	**M**	**SD**	**n**	**M**	**SD**
1	35	18.54	7.83	36	20.0	6.9
2	22	18.73	5.45	32	23.1	5.3
3	29	20.41	7.35	35	23.9	4.3
4	24	19.07	6.83	37	22.3	5.9
5				25	23.7	5.8

Table 4-2. Descriptive statistics for PSVT:R data collected from Bachelor of Electrical engineering years 1 to 4 and Bachelor of Architecture years 1 to 5, DIT, 2013-14.

In the case of the architecture students the result of the ANOVA was significant for both the PSVT:R scores ($F(4,164) = 2.689$, $p = .033$) and the MCT scores ($F(4, 153) = 5.931$, $p = .000$). As shown in Table 4-1 and Table 4-2, this result is due to the difference between the samples from years 1 and 2 with no differences between years 2 to 5. Assuming these samples come from the same population it can be argued the change is attributed to the programme of study, i.e. first year architecture leads to improvement in spatial ability as measured by these tests. Alternatively, it can also be argued that no spatial ability development occurs and the result is instead explained by disproportionately lower retention into second year of weak visualizers leading to an increased average in the second year sample. It's not possible to say where the truth lies and it may be somewhere in between these two arguments or there could be yet another reason. What is evident is that first year electrical engineering and architecture students are very similar in spatial ability but from second year onwards architecture students have stronger spatial skills on average.

Grades for all courses on the Bachelor of Electrical Engineering were also obtained in order to measure the relationship between spatial ability and academic performance in individual courses or modules and in each year. A semester in the electrical engineering programme typically contains six modules, each worth 5 ECTS. Exceptions include some 10 ECTS modules in the first year and a 15 ECTS project module in the final year. In addition, there are optional modules in third and fourth year which the students can select. In all, grade data were collected for 44 modules across the four years and were used to calculate an average grade for each year for each student.

	PSVT:R			MCT		
Average	**n**	**r**	**p**	**n**	**r**	**p**
Year 1	23	.305	.156	15	.006	.983
Year 2	22	.216	.333	13	.332	.267
Year 3	29	.398	.032*	33	.277	.119
Year 4	22	.310	.161	14	.338	.236
All 4 years	96	.287	.005**	75	.261	.024*

Table 4-3. Correlation between year averages and two spatial tests, PSVT:R and MCT, for the Bachelor of Electrical engineering, 2013-14

A series of Pearson bivariate correlation tests were then conducted between each spatial test score and grades in each module and year averages across all four years of the programme. Table 4-3 contains the results for each year average and the combined year averages. Of the 44 modules whose grades were examined in conjunction with the spatial tests, those that revealed a significant correlation are shown in Table 4-4.

		PSVT:R			MCT		
Year	**Course**	**n**	**r**	**p**	**n**	**r**	**p**
1	Computer programming I	33	.696	.000**	22	.538	.010**
1	Computer programming II	23	.479	.021*	15	.317	.250
3	Digital communications	17	.758	.000**	18	.716	.001**
3	Power electronics	9	.745	.021*	12	.584	.046*
3	Digital signal processing	29	.503	.005**	33	.430	.012*
4	Final year project	23	.430	.040*	24	.340	.235

Table 4-4. Correlations that passed a significance test of $p < .05$ between module or course grades and two spatial tests, PSVT:R and MCT, for the Bachelor of Electrical engineering, 2013-14

In terms of grade point average across all four years, approximately 8 % of the variation is shared with spatial ability as measured by the PSVT:R. It appears that this small but significant correlation is reflected by the relatively small number of modules, 6 out of 44, that revealed a correlation between grades and spatial ability. Small differences can be enough to change the classification of an award in the final year which can be a life changing event in terms of employment and post graduate study prospects for a graduate. With all other things being equal, this raised the possibility of cases where a student's position at either side of a threshold is determined by his/her spatial ability.

4.2 Spatial ability and academic measures among US engineering students

First year engineering at OSU represents a different context compared to that at DIT. Relatively larger numbers of students are enrolled at OSU. They are predominantly from Ohio and have matriculated from a very different secondary/high school learning environment compared to their peers enrolling in engineering at DIT. Cultural differences are found both within and without the educational setting. Engineering students at DIT are limited in their choice of subjects compared to their counterparts at OSU. For example, all students on the common first year engineering programme at DIT take the same modules – six per semester - and, in the case of mathematics, all take the same math course. Students typically proceed to level 8 programmes immediately after completing secondary or high school and entry to the programme depends on their school grades. In contrast, first year engineering students at OSU typically take four courses but this can vary as can the courses they take. All typically take a math course, with some variation on type of course, an engineering course that involves programming and microprocessors, a general education course, e.g. English, and another course of their choosing from a required list. Regardless of these differences, the aims and goals of each engineering programme are very similar and both programmes are accredited by their respective professional bodies, who, in turn, share much common ground with respect to programme outcomes in engineering education. First year engineering at OSU provided a different case study to DIT with regard to the role of spatial ability in the curriculum which presented an opportunity to learn more about the role of spatial ability in the engineering curriculum.

As part of the first year engineering orientation programme at OSU, incoming students must complete online assessments of spatial and mathematical abilities. These data are used to place students in a math course that is best suited to their abilities and to also identify those who have weak spatial visualisation skills. Those categorised as weak visualizers are advised to take a spatial skills training course that is designed to teach skills related to mental transformation between 2 and 3 dimensional drawings in the style of isometric sketches, orthographic projections, coded plans and sectional drawings (Sorby, 2009). Therefore, data collected during engineering orientation at OSU presented an opportunity to examine the relationship between spatial and math abilities and how together they predict performance in engineering education.

The research questions addressed in this section include:

1. Do spatial skills predict success on a mathematics placement test routinely administered at the university for first-year engineering students?
2. What is the nature or characteristics of the math placement test questions that reveal the biggest differences in spatial ability?
3. Is there a gender gap in MPT scores and, if so, to what extent is this explained by the gender gap in spatial ability?
4. To what extent are course grades related to spatial ability?

Research Design

Data were collected for this part of the study during the summer of 2016 using two instruments – the PSVT:R and the Mathematics Placement Test (MPT, unpublished) developed by the Department of Mathematics at OSU. These tests were administered online as part of the OSU freshman orientation programme and described in more detail below. Approval for the study was obtained from the Institutional Review Board at OSU. At the time of taking the PSVT:R test, participants were provided with an IRB approved document containing information about this study and asked to consent to participate by allowing their responses to the tests to be included in the data set. A total of n = 1053 out of approximately 2,400 students enrolled in first year engineering provided informed consent and only their data is presented and analysed here.

The MPT was developed by the Department of Mathematics at OSU and is not publicly available. It consists of 25 questions, each of which has 5 variations that differ only in numbers used and, therefore, all 5 versions can be considered equivalent. Reliability was measured as part of this study using Cronbach's α and found to be equal to .79 with none of the 25 items resulting in a higher value of α if removed. Therefore, the MPT is considered to have a high reliability. The topics included on the test are basic algebra, fractions, logarithm rules, exponent algebra, function notation, complex number arithmetic, inequalities, geometry and trigonometric rules and functions. One word algebra problem is included on the test and all of the questions are in multiple choice format.

Data collected from these participants also included scores on the following three US college entrance tests: the Scholastic Aptitude Test (SAT) Math and the American College Testing (ACT) Math tests, and the ACT Science Reasoning test (SCIRE). Grade point average (GPA) scores gained by the sample in the autumn semester, 2016, were also collected. Hence, the data set consisted of spatial ability, gender and five measures of student ability.

The ACT and SAT Math tests differ in that the SAT contains more logical reasoning and the ACT contains more advanced mathematics such as trigonometry. There is a greater time pressure in the ACT as the SAT allows 30 % more time per question, hence, speed is rewarded to a greater extent in the ACT. (https://greentestprep.com/resources/sat-prep/act-vs-sat/sat-act-differences/)

Both the PSVT:R and the MPT were administered as part of the orientation process through the OSU learning management system in the summer of 2016 immediately prior to enrolment. Participants took the tests online at a time and place of their choosing. Instructions were provided online for each test. Times allowed for the PSVT:R and MPT were, respectively, 25 and 75 minutes. Sample problems were available for the MPT and the PSVT:R test created by Guay (1976) contains two sample questions at the start.

The distributions of both the MPT and PSVT:R data were first examined and found to be skewed at the upper end of their ranges, particularly so for the PSVT:R (Figure 4-1). This is sometimes described as a ceiling effect with a large number of participants getting very high or maximum scores on the tests. This was more pronounced for the male compared to female cohort (skewness = -.930 for males vs. -.418 for females). However, since the sample size is large (n = 1053), the assumption of normality was presumed to hold in this case.

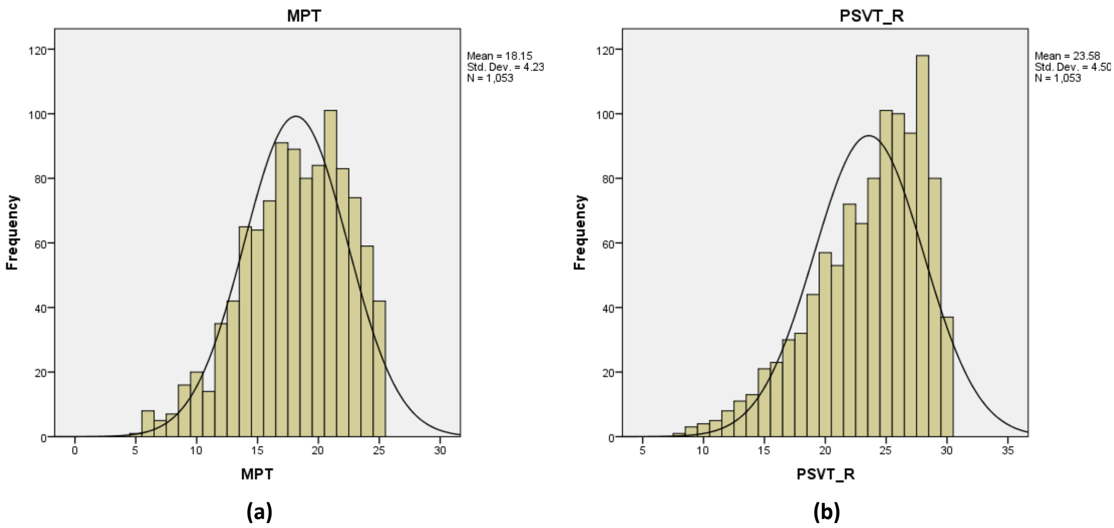


Figure 4-1. Distribution of (a) MPT and (b) PSVT:R scores for the OSU first year engineering sample.

Statistical analysis consisted of grouping the sample in two ways - by gender (male and female) and by spatial ability level (weak and strong) – and, for each grouping, comparing the scores of each of the five measures using an independent samples t-test to measure the significance of the difference and a Cohen's d effect size to measure the size of the difference. The interaction between gender and spatial ability level was also evaluated to determine whether the relationship between spatial ability and each measure depended on gender.

Results & Analysis

Descriptive statistics for the results from the MPT and PSVT:R tests are shown in Table 4-5 where they are presented for the entire sample and also grouped by gender. Included are the numbers of weak and strong visualizers in each group.

	All			**Male**			**Female**		
	n	**M**	**S.D.**	**n**	**M**	**S.D.**	**n**	**M**	**S.D.**
MPT	1053	18.15	4.234	781	18.45	4.208	272	17.29	4.198
PSVT:R	1053	23.58	4.506	781	24.28	4.155	272	21.55	4.855
Weak	151 (17 %)				79 (10 %)		72 (26 %)		
Strong	902 (83 %)				702 (90 %)		200 (74 %)		

Table 4-5. Descriptive data for the OSU freshman engineering sample presented for the full set and grouped by gender

These results show that 151 participants were categorised as weak visualizers and, therefore, eligible to take the spatial skills course offered in semester 1 at OSU. Performance on all of the tests are presented below and grouped first by gender (Table 4-6) and then by spatial ability level (Table 4-7) with differences in means of each group measured for significance by an independent t-test and for magnitude by Cohen's d effect size calculation ($d = (M_2-M_1)/\surd(SD_1^2+SD_2^2)/2$):

Test	**Male**			**Female**			**t-test**	**Sig (2-tailed)**	**Cohen's d (Size)**
	n	**M**	**SD**	**n**	**M**	**SD**			
MPT	781	18.45	4.208	272	17.29	4.198	3.906	.000	0.28 (Medium)
PSVT:R	781	24.28	4.155	272	21.55	4.855	8.931	.000	0.61 (Large)
GPA	777	3.07958	.738093	271	3.07368	.586109	.119	.905	0.01 (Small)
SAT Math	237	695.49	63.549	76	648.55	60.876	5.659	.000	0.76 (Large)
ACT Math	703	30.91	2.876	257	29.64	2.776	6.106	.000	0.45 (Medium)
ACT SCIRE	703	30.99	3.417	257	29.73	3.548	5.012	.000	0.37 (Medium)

Table 4-6. Differences in performance by gender on MPT, PSVT:R, GPA, SAT and ACT.

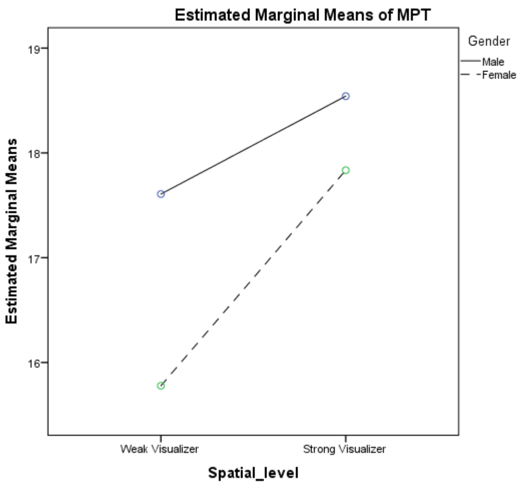
Test		Weak Visualizer			Strong Visualizer		t-test	Sig (2-tailed)	Cohen's d (Size)
	n	M	SD	n	M	SD			
MPT	151	16.74	4.29	902	18.38	4.18	-4.471	.000	0.39 (Medium)
PSVT:R	151	15.36	2.42	902	24.95	N/A			
GPA	148	2.84	0.67	900	3.12	0.70	-4.478	.000	0.41 (Medium)
SAT Math	39	664.10	62.69	274	686.93	66.04	-2.032	.043	0.36 (Medium)
ACT Math	134	28.92	2.58	826	30.84	2.87	-7.296	.000	0.71 (Large)
ACT SCIRE	134	29.22	3.45	826	30.89	3.45	-5.206	.000	0.49 (Medium)

Table 4-7. Differences in performance by spatial ability level on MPT, PSVT:R, GPA, SAT and ACT.

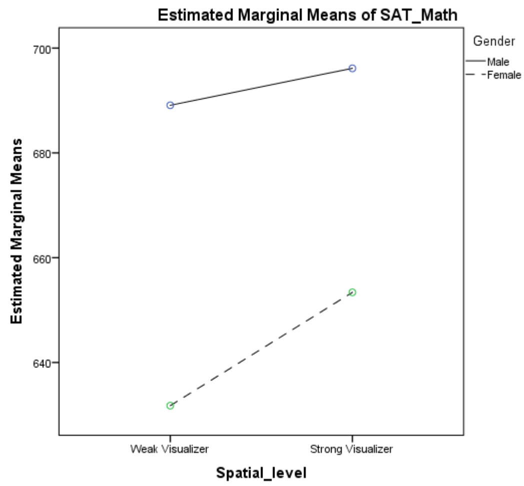
A two-way repeated measures ANOVA was conducted to measure the interaction of spatial ability level (weak or strong) with gender (male or female) on the different measures of academic performance collected from this sample (MPT, GPA, SAT, ACT, SCIRE) with results presented in Table 4 and presented graphically in Figure 4-2.

		Gender		Spatial ability		Gender x Spatial	
Variable	n	F	p	F	p	F	p
MPT	1053	11.180	.001	15.550	.000	2.194	.139
SAT Math	313	20.088	.000	1.649	.200	.426	.514
ACT Math	960	17.045	.000	38.505	.000	.503	.479
ACT SCIRE	960	11.208	.001	18.140	.000	.128	.720
GPA	1048	1.043	.307	19.717	.000	.335	.563

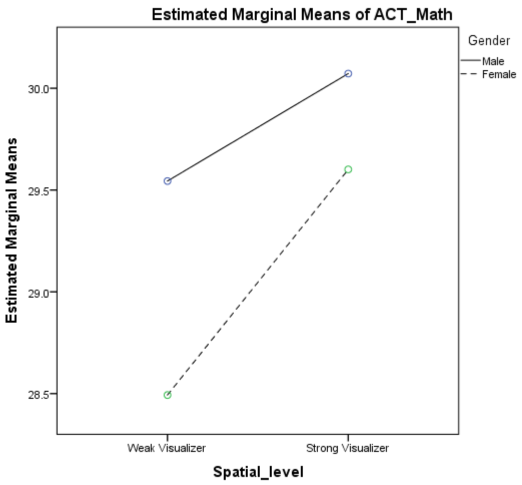
Table 4-8. Results of two-way repeated measures ANOVA to check for interaction between gender and spatial ability.



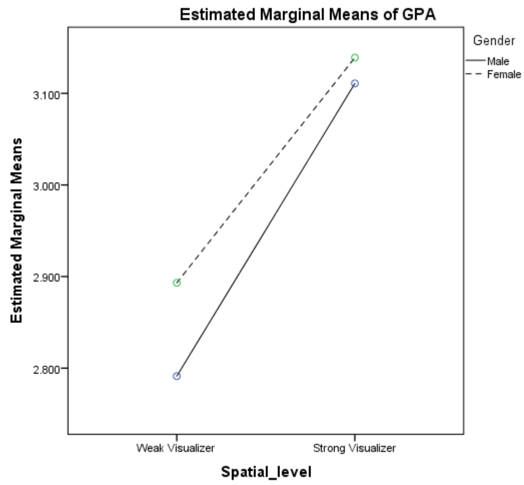
(a)



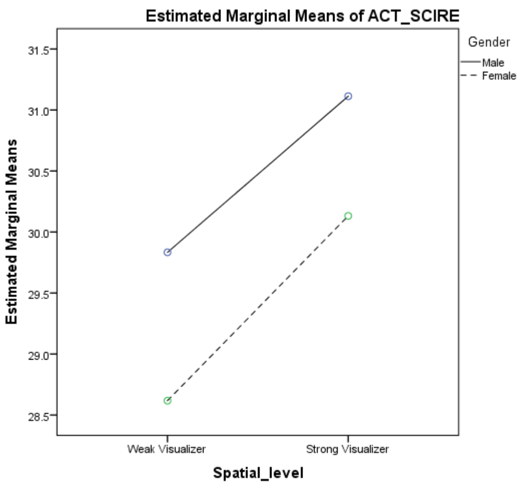
(b)



(c)



(d)



(e)

Figure 4-2. Graph of the interaction of gender and spatial ability level on (a) MPT, (b) GPA, (c) SAT Math, (d) ACT Math and (e) ACT SCIRE.

Finally, a correlation matrix is presented in Table 4-9 to show the extent to which each of the measures correlate with each other and with spatial ability based on the full data set, male and female.

	n	2	3	4	5	6
1. MPT	1053	.207**	.369**	.561**	.576**	.386**
2. PSVT_R	1053		.176**	.285**	.323**	.242**
3. GPA	1048			.264**	.413**	.302**
4. SAT_Math	313				.664**	.480**
5. ACT_Math	960					.506**
6. ACT_SCIRE	960					

** significant at $p < .01$.

Table 4-9. Correlation matrix for all variables measured.

The mean PSVT:R scores measured in this study for the entire sample, and separately for males and females are very similar to those reported for another group of US students (Sorby, Casey, Veurink, & Dulaney, 2013). The sample is also consistent with the internationally observed difference in spatial ability in favour of males (Lippa et al., 2010). In this case males scored significantly higher on the PSVT:R than females with the difference equal to 2.72 points (out of 30) and measured using Cohen's d to be a large effect size ($d = .61$, $p < .01$). Of the 272 females in the sample, 72, or 26 % of the females, were categorised as weak visualizers. In contrast, 79 of the 781, or 11 % of the males, were put in the same category. Hence, the weak visualizer cohort is 48 % female while the entire sample is 26 % female.

With regard to the first question, the extent to which spatial skills predict success on the MPT, the correlation between the two variables, although significant, is not very large ($r(1050) = .207$, $p < .01$). This result means that only a small amount of variation, 4.3 %, is shared between the PSVT:R and MPT scores. Hence, spatial skills, at least when measured by the PSVT:R, are not a strong predictor of success on the MPT. With regard to the other two math measures, ACT Math and the SAT Math, significant correlations were also measured but varied in magnitude. Ranked from smallest to largest correlation with the PSVT:R, the order is $r_{PSVT:R\text{-}MPT}(1050) = .207$, $r_{PSVT:R\text{-}SAT}(311) = .285$, $r_{PSVT:R\text{-}ACT}(958) = .323$, all $p < .01$. A bigger gap between weak and strong visualizers is revealed by ACT Math compared with SAT Math and MPT. No significant difference in PSVT:R scores was measured between those who took the SAT Math (M=23.87, S.D.=4.457) and those who took the ACT Math (M=23.63, S.D.=4.445) tests.

The second question relates to the issue of interaction between gender and spatial ability: is the relationship between spatial and math abilities different for males and females in this sample? The interaction between gender and spatial ability level was determined using a two-way repeated measures ANOVA, with each math measure separately entered as the dependent variable, which indicates there was no interaction between these variables on all three math measures (Table 4-8). The extent to which they interact is also shown in Figure 4-2 (a), (b) and (c) which illustrates males outperforming females in both spatial ability categories. In the case of the MPT, there is a noticeably larger difference between female weak and strong visualizers compared with male weak and strong visualizers. The correlations between the PSVT:R and MPT were measured separately for each gender to be $r_{male}(779) = .166$ and $r_{female}(270) = .228$, both $p < .01$. Although there is no crossover on the two lines in Figure 4-2 (a) and the interaction between gender and spatial ability level was found to be not significant, there is a difference in magnitude of the correlations between the two measures for each gender with the correlation being higher for females. This leads to the female weak visualizer being ranked lowest in MPT scores (M = 15.78), followed by the male weak visualizer who is very close in MPT to the female strong visualizer (17.61 and 17.84, respectively) and highest is the male strong visualizer cohort with an average MPT of 18.54 (Table 4-10). Being a female weak visualizer carries the highest risk of being ranked poorly on the MPT.

Spatial ability	Male			Female			Δ MPT	t-test	Sig (2-tailed)	Cohen's d (Size)
	n	M	SD	n	M	SD				
Weak	79	17.61	4.307	72	15.78	4.088	1.83	2.671	.008	0.44 (Medium)
Strong	702	18.54	4.190	200	17.84	4.112	0.7	2.112	.035	0.17 (Small)

Table 4-10. Comparison of MPT scores grouped by visualization category for male and female participants.

The MPT is used to place students in an appropriate math course in first year engineering at OSU. Those scoring 8 points or lower are enrolled in algebra; scores of 9 to 12 lead to pre-calculus enrolment; and scores of 13 or higher results in the student being enrolled in calculus provided his/her high school grades are satisfactory and, if not, the student could be enrolled in pre-calculus. The data were next checked to determine how each group – male and female weak and strong visualizers – were distributed across these MPT categories. These results are shown in Table 4-11.

MPT_level	N	Mean	Std. Deviation	n weak male	n strong male	n weak female	n strong female
MPT <= 8	21	22.52	4.844	2 (0.3%)	13 (1.7%)	3 (1.1%)	3 (1.1%)
9 <= MPT <= 12	85	22.13	4.649	6 (0.8%)	45 (5.8%)	14 (5.1%)	20 (7.4%)
13 <= MPT <= 17	335	22.82	4.584	27 (3.5%)	211 (27%)	28 (10.3%)	69 (25.4%)
MPT >= 18	612	24.23	4.321	44 (5.6%)	433 (55.4%)	27 (9.9%)	108 (39.7%)
Total	1053	23.58	4.506	79 (10.1%)	702 (89.9%)	72 (26.5%)	200 (73.5%)

Table 4-11. Distribution of sample by gender and spatial ability level across the four MPT levels. (Percentages based on total amount of each gender).

Although the correlation between the two measures, MPT and PSVT:R, is not large, the results in Table 4-11 show a difference in the number of students placed in each MPT category when grouped by gender and spatial ability level. Since the occupancies of the MPT ≤ 8 categories are very small they were grouped with the next category for analysis. Hence, 23 % of the female weak visualizers have a MPT ≤ 12 which is much higher than the equivalent percentages for the other groups. The distribution of male weak visualizers across the MPT categories is very similar to that of the female strong visualizers. As shown in Table 4-11, female weak visualizers have the greatest proportional representation in the algebra class (MPT ≤ 12). This class is overrepresented by women who comprise 40 % of the class versus 26 % of the sample. The small correlation suggests the cognitive processes involved in answering questions on the MPT and the PSVT:R overlap to a small extent only but that, particularly for females, being classified as a weak visualizer predicts a higher risk of poor performance on the MPT. It has been suggested by others that the relationship between spatial and math skills may not be the same for males and females because spatial ability plays a more important role in math performance for females (Halpern et al., 2007). These data do not contradict this assertion.

The third question relates to the relationship between performance in STEM education and spatial ability. Since it has been shown that measures of spatial ability and mathematics can explain more variance in academic achievement scores than mathematics alone it was decided to test this hypothesis by examining results by gender. GPA based on the first semester of first year was used as a measure of academic performance. During the first semester, there is much variation in what courses are taken and contribute to GPA but the majority of first year engineering students at OSU are likely to take a course in math, general education and chemistry and one of four engineering courses. What contributes to GPA is, therefore, quite varied but it was the only measure of academic performance in semester 1 that was available. Using an independent t-test, no significant difference in GPA by gender was found but a difference was found when the sample was grouped by spatial ability ($d = .41$, $p < .001$, Table 4-7). The relationship between spatial ability and GPA was very similar for both genders (Figure 4-2 (d)) and no interaction between spatial ability level and gender was observed for GPA (Table 4-8). The correlation between the PSVT:R and GPA was $r(775) = .177$, $p < .01$, for male and for female $r(269) = .208$, $p < .01$, i.e. very similar. The correlation is small – a change of 10 points in the PSVT:R score, which is large, is approximately equivalent to a change of 0.3 in GPA.

While the correlation and effect size are small, a small change in GPA can have a big effect, particularly at borders such as GPA = 2.0 and GPA = 3.0 below which a student can be placed on academic probation, depending on the institute. Indeed, as shown in Figure 4-2, strong visualizers are on the right side of the GPA = 3.0 border. The sample was grouped into three GPA levels – GPA < 2.0, 2.0 ≤ GPA 3.0 and GPA ≥ 3.0 and the number of male and female weak and strong visualizers in each category was counted with the results are presented in Table 4-12.

These results (Table 4-12) are slightly more favourable from the female perspective with proportionally greater representation of men in the lowest GPA category. The pattern in the highest GPA category is very similar for male and female weak visualizers and likewise for male and female strong visualizers. This is consistent with the comparison made in Table 4-6 above, i.e. when compared by gender there was no difference in mean GPA between male and female.

GPA level	n	Mean	Std. Deviation	n weak male	n strong male	n weak female	n strong female
GPA < 2	83	23.04	4.964	8 (1%)	60 (7.7%)	5 (1.8%)	10 (3.7%)
2.0 <= GPA < 3	317	22.54	4.710	33 (4.2%)	188 (24.2%)	35 (12.9%)	61 (22.5%)
3.0 <= GPA < 4	648	24.19	4.235	35 (4.5%)	453 (58.3%)	32 (11.8%)	128 (47.2%)
Total	1048	23.60	4.504	76 (9.8%)	701 (90.2%)	72 (26.6%)	199 (73.4%)

Table 4-12. Distribution of sample by gender and spatial ability level across three GPA levels. (Percentages based on total amount of each gender).

The last question related to the nature of the questions on the MPT in terms of their relationship with spatial ability. Which questions explained the correlation between the two measures and why did these questions draw more than others on spatial thinking? An example of the MPT is provided in the Appendix D. For each question in turn, the sample was grouped by correct or incorrect response to the question and the spatial test scores of these two groups were then compared using an independent t-test. Thirteen of the 25 problems revealed no difference in spatial ability, i.e. the mean PSVT:R score of the correct and incorrect groups were equal. These were very much plug and chug type operations in which an equation was provided and a procedure had to be followed. These questions tested core competencies in mathematics at a procedural rather than a problem solving level; a problem representation step was not required. For 12 of the 25 questions, the sample was divided to reveal significant differences in PSVT:R but in many cases the difference was small. Only two questions had an effect size greater than 0.4 which equated to difference in mean PSVT:R scores between the correct and incorrect groups of 1.8 points. One of these questions was an algebra word problem related to mixing two components that required the translation of a problem statement into equations which are then solved for two unknowns. The other required the comprehension of an inverse relationship between two variables. Examples of these two questions are provided below:

(i) How many pounds of M&Ms that cost $1.60 per pound must be mixed with 7 pounds of Smarties that cost $2.20 per pound to make a mixture of sweets that costs $2.00 per pound?

(ii) Let x vary inversely as y. When x is 21, y is 4. When x is 7, y is …

Q	Math topics	Schema	Correct			Incorrect			t-test	p	Cohen's d
		provided	n	M	SD	n	M	SD			(Size)
1	Fractions	Yes	960	23.61	4.50	93	23.20	4.55	-.838	.402	0.1 (Small)
2	Logs, exponents, prime factors	Step 1 No Step 2 Yes	890	23.50	4.55	163	24.01	4.25	1.319	.188	0.12 (Small)
3	Algebra, exponents	Yes	656	23.66	4.55	397	23.45	4.43	-.728	.467	0.05 (Small)
4	Function notation, domain definition, algebra	Yes	1038	23.61	4.49	15	21.27	5.23	-2.004	.045*	0.49 (Medium)
5	Complex number arithmetic	Yes	663	24.03	4.29	390	22.81	4.76	-4.292	.000**	0.27 (Medium)
6	Algebra, exponents	Yes	1004	23.63	4.50	49	22.53	4.64	-1.668	.096	0.25 (Medium)
7	Algebra, exponents	Yes	689	23.97	4.44	364	22.83	4.54	-3.946	.000**	0.26 (Medium)
8	Inverse function	Yes	614	24.35	4.10	439	22.50	4.83	-6.709	.000**	0.42 (Medium)
9	Algebra, number line, inequalities	Yes	808	23.77	4.41	245	22.96	4.76	-2.461	.014*	0.18 (Small)
10	Handling exponents	Yes	831	23.78	4.54	222	22.83	4.33	-2.782	.006**	0.22 (Medium)
11	Algebra, n equations for n unknowns	No	626	24.31	4.24	427	22.50	4.67	-6.547	.000**	0.41 (Medium)
12	Algebra, number line, inequalities	Yes	771	23.85	4.53	282	22.84	4.36	-3.213	.001**	0.23 (Medium)
13	Complex number arithmetic	Yes	886	23.78	4.41	167	22.51	4.86	-3.360	.001**	0.28 (Medium)
14	Circle equation, algebra	Yes	862	23.83	4.46	191	22.45	4.54	-3.867	.000**	0.31 (Medium)
15	Angular velocity, convert rpm to rads	No	893	23.74	4.44	160	22.68	4.78	-2.743	.006**	0.23 (Medium)
16	Trigonometry, sin defn, angles	Yes	944	23.67	4.46	109	22.81	4.88	-1.889	.059	0.19 (Small)
17	Trigonometry, cot defn	Yes	941	23.64	4.48	112	23.04	4.71	-1.249	.178	0.14 (Small)
18	Trigonometry, sin defn, angles	Yes	842	23.70	4.45	211	23.09	4.71	-1.745	.081	0.14 (Small)
19	Trigonometry, meaning of exact, exact values for angles 30 and 45	Not clearly	650	23.98	4.49	403	22.93	4.46	-3.724	.000**	0.24 (Medium)
20	Trigonometry	Yes	540	23.86	4.45	513	23.28	4.55	-2.079	.038*	0.13 (Small)
21	Trigonometry, angles	Yes	707	23.71	4.38	346	23.32	4.73	-1.327	.185	0.09 (Small)
22	Polygon, geometry	Not clearly	446	24.32	4.33	607	23.03	4.56	-4.626	.000**	0.3 (Medium)
23	Trigonometry definitions	Yes	614	23.69	4.58	439	23.42	4.41	-.983	.326	0.07 (Small)
24	Geometry, trigonometry	Yes	682	23.75	4.52	371	23.27	4.47	-1.642	.101	0.11 (Small)
25	Trigonometry	Not clearly	553	23.97	4.50	500	23.14	4.48	-2.999	.003**	0.19 (Small)

Table 4-13. For each MPT problem, math topics and schema, PSVT:R scores for sample grouped by correct and incorrect on each problem and differences in these scores.

4.3 Spatial ability and comprehension of electric circuits knowledge

Electric circuits is a core subject in electrical engineering that is typically introduced in the first few semesters of study. Learning outcomes relate to the comprehension of several aspects of simple direct current (DC) electric circuits. These include concepts of energy, voltage/potential difference and current and also include the ability to create and interpret formal circuit diagrams which requires knowledge of both symbols and the application of the concepts. Since previous studies (e.g. Kozhevnikov et al., 2007) have shown a link between spatial ability and ability to reason about concepts from Newtonian mechanics it was decided to examine the potential relationship between reasoning about electric circuits, a key topic in electrical engineering, and spatial ability. The research questions addressed in this part of the study are:

1. What are the spatial ability characteristics of freshman engineering students in DIT?
2. To what extent does a core subject of electrical engineering, electric circuits, depend on spatial thinking?

Data collection

At the beginning of the academic year (2014/15) the first year of the electrical engineering programme was replaced with a general entry, common first year programme taken by all Bachelor of engineering candidates. Two spatial skills tests – the MCT and the PSVT:R - were administered to this group in the first few weeks of the semester (September 2014). A module on electrical systems was delivered during the first semester with a closed book written examination worth 75 % of the module grade taken by the entire class (n = 166) at the end of the semester. Laboratory work accounted for the remaining 25 %. The lecturer for this module administered the DIRECT electric circuits concept test twice - early in the semester (September 2014) and again towards the end of the semester (late November 2014). Both the spatial and DIRECT tests were administered in class with variability in attendance and response rates to the tests. These data were collected again the following year with the next class of first year engineering students (2015/16) with both the spatial and DIRECT tests administered at similar points in the semester. Additional spatial and DIRECT data were collected for the third year electrical engineering class in October 2015 and in Australia these tests were administered to two classes at the University of New South Wales (UNSW).

Presented in Table 4-14 are the samples from whom data were collected, the type of spatial test administered and descriptive statistics for each measure, spatial ability and DIRECT.

	Sample	Data collected	Spatial test	n	M	SD	DIRECT n	M	SD
1	DIT 1st year common engineering	September 2014	PSVT:R	131	21.67	6.13	133	8.61	3.16
1	DIT 1st year common engineering	September 2014	MCT	125	11.88	4.76			
1	DIT 1st year common engineering	November 2014	PSVT:R	107	23.42	5.50	103	13.57	3.97
2	DIT 3rd year electrical engineering	October 2015	MCT	27	10.41	5.34	27	12.33	3.84
2	DIT 3rd year electrical engineering	October 2015	MRT-A	27	10.63	5.58			
3	DIT 1st year common engineering	October 2015	PSVT:R Redrawn	84	21.69	5.40	90	14.33	3.56
4	UNSW Undergraduate	February 2016	MCT	31	15.52	5.85	31	19.00	4.32
5	UNSW Postgraduate	February 2016	MCT	66	13.64	5.02	66	19.08	5.14

Table 4-14. List of sample from which DIRECT and spatial ability data have been collected with descriptive statistics.

For each sample, a Pearson correlation was calculated between the spatial test and DIRECT with these results provided in Table 4-15.

Sample	Data collected	Spatial test	n	r_{DIRECT}	r_A[1]	r_B[1]	r_C[1]	r_D[1]
1, DIT 1st year common eng.	September 2014	PSVT:R	112	.218*	.356**	.075	-.041	-.012
1, DIT 1st year common eng.	September 2014	MCT	114	.233*	.435**	-.049	-.002	-.020
1, DIT 1st year common eng.	November 2014	PSVT:R	80	.459**	.458**	.247*	.069	.326**
2, DIT 3rd year electrical eng.	October 2015	MCT	27	.492**	.527**	.298	.001	.320
2, DIT 3rd year electrical eng.	October 2015	MRT-A	27	.264	.505**	-.112	.138	-.081
3, DIT 1st year common eng.	October 2015	PSVT:R Redrawn	84	.268*	.360**	.078	.047	.093
4, UNSW UG electrical eng.	February 2016	MCT	31	.253	.468**	N/A	N/A	N/A
5, UNSW PG electrical eng.	February 2016	MCT	66	.489**	.488**	N/A	N/A	N/A

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).
[1]. A = Physical aspects of DC circuits, B = Energy, C = Current, D = Potential Difference or Voltage

Table 4-15. Correlations between spatial ability and DIRECT (total score and Group A, B, C and D) for each sample in the study.

As shown in Table 4-15, a sizeable and significant correlation was found between spatial ability and the DIRECT test which was large in some cases, e.g. $r(26) = .492$, $p < .01$ for 3rd year electrical engineering students at DIT. However, it is not consistent across samples – both magnitude and significance vary greatly. There are several possible reasons for this. The first relates to variation in prior knowledge of DC electric circuits – some samples may have taken more classes in the subject than others. For sample 1, for example, the first set of DIRECT data were collected at the beginning of the academic year after only a couple of weeks of formal instruction in DC electric circuits principles. The test was administered at this point by the module instructor in order to identify at an early stage in the module what misconceptions should be targeted during the semester. Given that many had not studied the topic before, it could be argued that many responses to DIRECT at this stage are random selections of the multiple choice options and, hence, a possible reason for the inconsistency. Indeed, an average score of 8.61 is only slightly higher than chance. However, when rows 1 and 2 in Table 4-15 are excluded, a variation in the correlations remains. Therefore, variation in prior knowledge is either not the reason or not the only reason for the differences in correlations.

The variation in the correlations could also be the result of the variety of spatial ability measurement used – PSVT:R, MCT and MRT-A – each measuring aspects of spatial ability that are different enough to affect correlations with the DIRECT test. However, noticeable differences in correlations are observed among the three DIRECT to PSVT:R correlations and the four DIRECT to MCT correlations so even when the same test is used differences at the correlation level are found. Therefore, variation in test type is not the only reason for the differences in correlation values.

Another explanation lies in the composition of the test itself. DIRECT contains 29 multiple choice questions that were designed by its authors to assess four separate conceptual areas – physical aspects of DC circuits (Group A), energy (Group B), current (Group C) and voltage (Group D). The questions that correspond to each area are outlined in their paper (Engelhardt & Beichner, 2004, Table I). To investigate the possibility that differences in correlations were connected to these groupings, the scores on DIRECT were separated into these four groups and individual correlations with spatial test scores were calculated which are presented in Table 4-15.

A much more consistent pattern emerged for the correlations with the sub-scores on DIRECT compared with the total score on DIRECT. The correlation between spatial ability and Group A questions is repeatedly significant at the $p < .01$ level while for the other three groups the correlation is variable but insignificant in most cases. Hence, the consistent aspect of the relationship between spatial ability and DIRECT is found in tasks related to physical aspects of circuits, Group A. Correlations with the other three groups are more inconsistent which contributed in no small way to the differences in correlations with the entire DIRECT test. The average value of the correlations between the spatial test and Group A on DIRECT is .45 and all are significant at $p < .01$. If the correlations measured early in the semester of year 1 (rows 1 and 7) are removed the average becomes .47 with a range of .36 to .53. On average, 23 % of the variation in scores on Group A of DIRECT are shared with a test of spatial ability and the highest correlations are observed when the MCT is used to measure spatial ability.

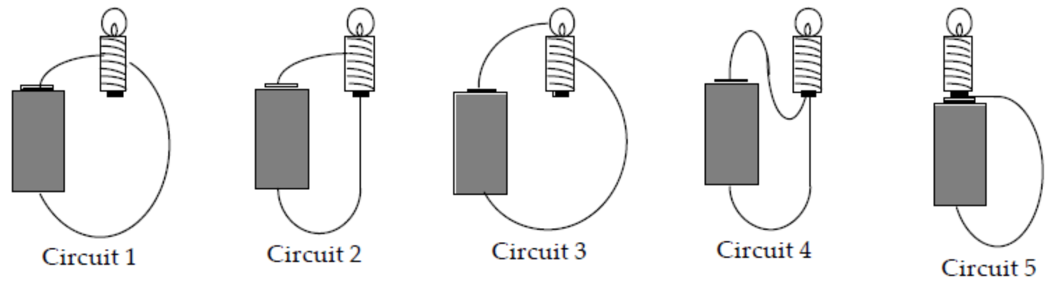
Objective	**Question No.**
(1) <u>Identify</u> and explain <u>a short circuit</u> (more current follows the path of lesser resistance).	10, 19, 27
(2) Understand the functional two-endedness of circuit elements (elements have two possible points with which to make a connection).	9, 18
(3) <u>Identify a complete circuit and understand the necessity of a complete circuit for current to flow</u> in the steady state (some charges are in motion but their velocities at any location are not changing and there is no accumulation of excess charge anywhere in the circuit).	
Objectives 1 to 3 combined.	27
(4) Apply the concept of resistance (the hindrance to the flow of charges in a circuit) including that resistance is a property of the object (geometry of object and type of material with which the object is composed) and that <u>in series the resistance increases as more elements are added and in parallel the resistance decreases as more elements are added</u>.	5, 14, 23
(5) Interpret pictures and diagrams of a variety of circuits including series, parallel, and combinations of the two.	4, 13, 22

Table 4-16. Objectives for the Group A questions on DIRECT (Engelhardt & Beichner, 2004, p. 100)

All 11 Group A questions share a similar format which consists of a short question, an image of a circuit with multiple answer choices provided for the participant to select (e.g. Figure 4-3). A copy of the DIRECT test is available in Appendix C. According to Engelhardt & Beichner (2004, p. 100), the circuits concepts assessed by Group A are outlined in Table 4-16. Underlined in Table 4-16 are arguably the primary aspects of circuits that are assessed by this group of questions. Five of the Group A questions require transformation between pictorial sketch and formal circuit diagram (questions 9, 13, 18, 22 and 27). Four require some rearrangement of a formal circuit diagram (questions 4, 5, 10 and 19) and two require reasoning about circuit properties when a switch is closed (questions 14 and 23).

The format of Group A questions is different to mathematical story problems in that they do not contain a story, they do contain an image, both mental transformation and reasoning are required and answer choices are provided. While the multiple choice question format can prompt representations of the problem and facilitate guess work, those who do attempt to answer the questions are faced with decision making that can be categorised as problem representation. As noted by Engelhardt & Beichner (2004, p. 100) in item 4 of Table 4-16, examples of representation within this subject area include ‘resistors in parallel’ and ‘resistors in series’. Once one of these representations is selected, the corresponding equation can then be recalled and used to determine the answer.

27) Will all the bulbs be the same brightness?



(A) Yes, because they all have the same type of circuit wiring.
(B) No, because only Circuit 2 will light.
(C) No, because only Circuits 4 and 5 will light.
(D) No, because only Circuits 1 and 4 will light.
(E) No, Circuit 3 will not light but Circuits 1, 2, 4, and 5 will.

Figure 4-3. Question 27 from DIRECT, Group A (Engelhardt & Beichner, 2004).

Since assessments of electric circuits learning outcomes also require math skills it was decided to compute and compare correlations between grades in the electric circuits module, the mathematics module, DIRECT and spatial ability using sample 1, the first year engineering class of 2014/15, for whom all of these data sets were available. As shown in Table 4-17, the correlation between the DIRECT test (all groups) and the math module was negligible and insignificant whereas the correlation between the math and circuits module is large and significant. This implies that DIRECT does not draw on math skills but does require a large element of spatial thinking (due to the Group A questions). In the circuits module, one appears to be rewarded more for math skills than for the concepts assessed by DIRECT.

Test[a]	PSVT:R Post n=79	Circuits n=102	Math n=102
DIRECT Post	.459**	.373**	.072
PSVT:R Post		.265**	.052
Circuits Module			.628**
Math module			

**. Correlation is significant at the 0.01 level (2-tailed)
*. Correlation is significant at the 0.05 level (2-tailed).

Table 4-17. Correlation matrix for variety of measures from DIT 1st year engineering 2014/15.

Grades in the circuits course are based on laboratory work, with a 25 % weighting, and a closed book final examination weighted at 75 %. Laboratory work is procedural in nature and does not require problem solving. The exam consisted of four questions, the first being a 40 item multiple choice short quiz and the remaining questions requiring students to write their own answers. The second question has two parts with the first containing circuits problem without instructions on how to solve it while the second part consisted of 5 short questions in which the candidate had to calculate voltage, current, resistance and power at particular points in the circuit. Advice was not given on how to do this but they were quite routine in nature. The first part of Question 3 (20 out of 30 marks available on the question) covered the same DC circuit material but advice was given to candidates to employ Thévenin's theorem in solving the question and the second part assessed material not included in DIRECT.

Grade data from the circuits module was found to have a significant relationship with the spatial test ($r(100) = .265$, $p < .01$). In order to investigate this relationship further, data consisting of students' solutions to the electric circuits end of semester closed book written examination were obtained from the course instructor for the 2015/16 class group. A short description of each question, the maximum mark, average score and standard deviation for each question and a correlation between scores on each question and the spatial ability test, the Revised PSVT:R in this case, are shown in Table 4-18.

Question	Description of Exam Question	Max	M	SD	r	p	n
1	40 MCQ, the first 22 assess DC circuit core competencies	40	21	6.1	-.001	.992	136
2 (a)	Calculate the value of one unknown resistor in a series parallel circuit. Circuit diagram provided but instructions on how to solve are not provided.	17	8.97	5.25	.249	.006**	119
2 (b)	For another circuit diagram, calculate values for voltage, current, resistance and power at particular points in the circuit.	13	4.10	2.73	.172	.062	119
3 (a)	20 marks based on circuit analysis problems related to simple DC circuits concerning resistance, current and voltage with instruction to use a schema (Thévenin's theorem)	20	4.8	5.1	-.017	.866	105
3 (b)	10 marks related to frequency, waveform and r.m.s. values, issues not addressed in DIRECT	10	3.7	3.1	.001	.993	105
4	A series of short questions related to procedural knowledge of a.c. circuit analysis	30	8.5	8.2	.157	.067	136
Total	Answer Question 1 and 2 others	100	40	15.8	.118	.172	136

**. Correlation is significant at the 0.01 level (2-tailed).

Table 4-18. Correlations between the Revised PSVT:R and scores on the final examination for the common 1st year engineering circuits module, DIT, 2015/16

As shown in Table 4-18, the spatial test scores share nothing in common with the scores on Question 1 and likewise for Questions 2(b), 3 and 4. Question 2 (a), however, has a small but significant correlation with the Revised PSVT:R scores. It is interesting to note that this question didn't provide the candidate with any direction on how to answer the question. The candidate had to decide what approach to take.

These data can be interpreted to show that tests of procedural knowledge, or core competencies, in electric circuits are independent of spatial thinking and that problem solving that requires a representation step correlates to a significant extent with spatial ability. However, this conclusion, particularly the latter part, is based on a very limited data set. A test that contained several problems similar to Question 2(a) on this exam, i.e. that require the participant to represent the problem, would be needed along with a set of questions that are more closely aligned with assessing the core competencies needed to solve the problems.

Conclusions

OSU students begin their engineering education with slightly higher levels of spatial ability than their peers at DIT with mean PSVT:R scores of 23.58 vs. 21.67 measured, respectively, from a sample at each institute. Possible explanations include the differences in the number of females in each sample, ethnicity, socioeconomic status, secondary or high school learning environment or any combination of these. Given the gender and, sometimes, ethnicity gap in spatial ability (Lippa et al., 2010) it is quite reasonable to expect a difference in PSVT:R scores were the DIT sample to contain significantly more women or more students from the Arabian Peninsula. Unfortunately, gender and ethnicity data were not available for the DIT sample. Anecdotally, however, the DIT sample is more likely to contain fewer women than OSU. Spatial ability has been shown to increase with socioeconomic status (Lippa et al., 2010; Wai et al., 2009) and while Ireland is a relatively affluent country it entered an economic crisis close to the time the Irish students in the DIT sample began their secondary school period. Data related to socioeconomic status for either sample was unavailable. Finally, there are differences in the secondary or high school curricula in each country and it is possible that greater opportunities for spatial ability development are available in the US but a comparison of the two schooling systems is not made here. For whatever reason, the OSU students gained higher scores on the PSVT:R that are very similar to measurements taken from another sample of US students (Sorby et al., 2013) and to the post test, end of semester scores collected from the DIT students.

A snap shot of spatial ability levels taken in one academic year from students on all four years of the Bachelor in Electrical Engineering, DIT, revealed no significant difference in spatial ability with time spent studying electrical engineering. In contrast, those enrolled in the Bachelor of

Architecture, DIT, revealed an increase in spatial ability during the first year of the programme but this could also be explained by poor retention of weak visualizers on the programme. While this could be explained by differences in the samples it is plausible that a student's spatial ability is not likely to be changed by studying electrical engineering in DIT. If true, then four years of higher level education in electrical engineering does not manifest in changes in spatial cognition as measured by the PSVT:R or MCT, a somewhat surprising thought.

DIRECT, an electric circuits concept inventory, was found to have a significant relationship with spatial ability not because of questions on DIRECT related to concepts of current, energy and voltage but because of questions that assessed the physical aspects of DC circuits. These circuits are figural objects that must be mentally transformed and rearranged and the questions also require some problem representation such as 'resistors in series' or 'resistors in parallel'. Correlations with DIRECT Group A were higher for the MCT than the PSVT:R possibly indicating a greater element of mental transformation than rotation when manipulating the circuit diagrams. This trend was repeated across several diverse samples and suggests that weak visualizers are more likely to make errors when working with circuit diagrams and be more challenged than strong visualizers when taking circuits courses. Compared to the correlation with DIRECT, that of spatial ability with the electric circuits exam was lower which implies the exam does not stress the spatial visualization and circuits reasoning skills required by DIRECT. Weak visualizers performed equally well as strong visualizers on the MCQ portion of the exam, for example. DIRECT revealed a large gap in the ability to comprehend physical aspects of DC circuits between weak and strong visualizers but grades in the corresponding subject indicate that learning outcomes are assessed in such a way that weak visualizers are not disadvantaged as much as they could, or possibly should, be.

The results highlight an apparent contradiction between lack of development in spatial ability in the electrical curriculum on the one hand and yet a significant correlation between spatial ability and a core component of electrical engineering on the other. The curriculum, however, contains many assessment components only some of which relate to circuits concepts and, in addition, if many assessments heavily reward rote learning then a relationship with spatial ability is less likely to emerge. The results do raise the possibility that if weak visualizers improved spatial ability they could become better electrical engineering students. Such a contradiction is less likely to exist for mathematical ability as its development is formally addressed in the curriculum by requiring engineering students to take several math courses spread across all years of the programme, i.e. studying engineering requires good mathematical ability and mathematical ability is developed by studying engineering.

In the OSU context, spatial ability contributed to a small extent only to the different measures of academic performance but, in GPA, these differences are sufficient to place a student on the right side of the GPA = 3.0 threshold. Female weak visualizers were more likely than others to be placed in the lowest MPT ranking: 23 % of them were found here compared with 8 to 12 % of the other categories. Differences in MPT when grouped by spatial ability were larger for females than males but not to the extent that there was significant interaction between the two attributes.

Correlations between spatial ability and aspects of STEM education can often be hidden by data collected from course grades and this is illustrated in both the DIT and OSU contexts. Spatial ability and electric circuits exam data from DIT yielded a correlation of $r(100) = .265$, $p < .01$, a much smaller amount of shared variation than that revealed by the correlation with DIRECT. Hence, there is an aspect of this subject area that is very strongly related to spatial ability but it is not revealed by course grades. Similarly, a very small correlation was observed between spatial ability and the MPT which masked a strong relationship with a small number of the problems on the test. This illustrates an aspect of the relationship between spatial ability and STEM education – it can often be hidden behind correlations with grades in subjects and/or GPA scores. These correlation can be so small as to discourage further investigation and lead to the erroneous conclusion that no relationship between spatial ability and this subject area exists. Course or module grades can be very broad in the range of skills and knowledge they assess or they may predominantly reward memorisation and, in either case, relationships with spatial ability can remain hidden.

Finally, the finding that was hidden amongst the MPT results was that spatial ability has a much more significant relationship with word story problems than tests of procedural knowledge. It was the mathematical story problem in the MPT that revealed the biggest difference in spatial ability between those who got each question correct and incorrect even though the mathematical procedures needed to solve this problem were much simpler than for the MPT questions where no difference in spatial ability emerged. Do these findings show that most of the problems on the MPT are not problems at all but assessments of math competencies that have been practiced repeatedly over time? Can these findings be taken to mean that differences in spatial ability are only revealed when the mathematical task is not procedural but is a problem in the true sense of the word, i.e. a novel situation where initial and goal states are clear but the path from one to the other is not? The MPT was not designed to answer these questions. First, the MPT contained only one story problem and 24 non-story questions and second, it was a multiple choice test so there were multiple answers were available to the students when solving the one story problem which may have assisted them in the problem solving process. Therefore, to address these questions a different set of math instruments was required to learn spatial ability is important in solving simple story problems in mathematics. This is the theme that is addressed in detail in the coming chapters.

Chapter 5 Spatial ability and success in solving algebra word problems

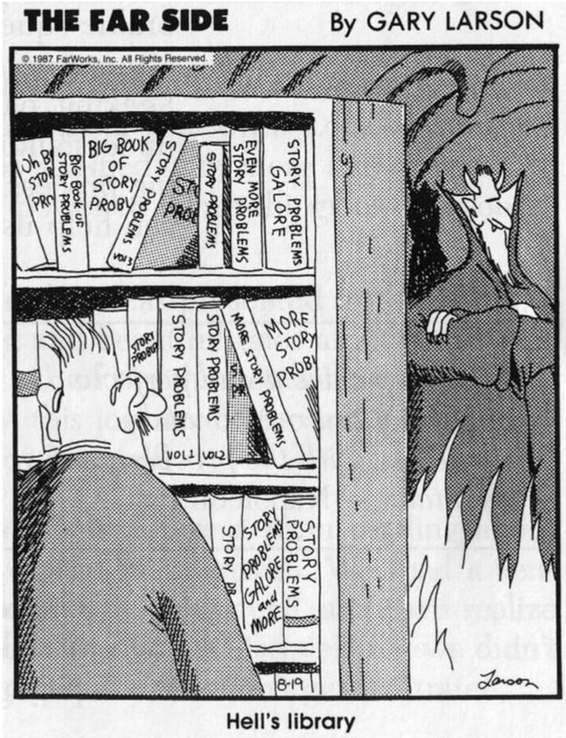


Figure 5-1. 'Hell's library' by Gary Larson. Reprinted with permission ("Andrews McMeel Syndication - Home," 2017)

Introduction

It is known that spatial ability correlates with success in non-routine problem solving activities in STEM education and, in the case of mathematics, with reasoning and problem solving tasks. The problem solving process has been conceptualised as having two steps - problem representation followed by problem solution – with these two steps requiring different types of thinking (Mayer, 1992). Problem solution requires the careful, algorithmic application of core math competencies that can be developed through practice and effort. Problem representation operates at a higher level of interpretation to the point that a slight rephrasing of problem statements can evoke different approaches to solving the problem (e.g. Coquin-Viennot & Moreau, 2003). It is in the problem representation step that the approach to solving the problem is decided. Success or failure hinges on this step. Is this why spatial ability correlates with measures of problem solving? Do problem representation and spatial ability share common ground at a cognitive level? Bringing together these two phenomena – the correlation between spatial ability and problem solving and the two step model of problem solving – leads to the following hypothesis:

> *Strong visualizers will outperform weak visualizers on measures of problem representation but will perform equally well as weak visualizers on measures of problem solution.*

In order to address this hypothesis, the separate measurement of spatial ability, problem representation and problem solution is required and, therefore, a measurement instrument is required for each variable. Of these three, selection of the spatial test is the most straightforward as many established options are available and reported in the literature and, of these, tests of mental rotation have yielded the highest correlations with math (Friedman, 1992). The PSVT:R is a mental rotation test that has been used extensively with engineering students (Maeda & Yoon, 2011; Sorby & Veurink, 2010b) and was used in this study.

More thought was required to determine how to separately measure problem representation and solution. One option is to create two tasks, one consisting of several problem representation challenges and the other several problem solution challenges, and administer each one. They should be paired, i.e. the solution would have to be matched with or belonging to the representation. The representation challenge would be phrased as a problem but not fully solved as the participant could be asked to stop after the representation has been made. This approach would work for measurement of the solution but measuring representation alone appeared to be difficult and a simpler approach was available which was to separately measure problem solving and problem solution. In the problem solving test, both representation and solution are measured as a participant solves the problem and navigates first through the representation step and then the solution step. In the second test, only the solution phase is measured by presenting a participant with a problem that has been represented and is now a mathematical task with instructions of how to complete it. Problem representation is, therefore, the difference between the two measures. This was the approach adopted in this study.

A set of problems was created to use as the problem solving instrument. For each problem, the core competency or competencies associated with it, such as solving simultaneous equations, were identified to create a separate set of ‘questions’. For example, in the ‘Lawn Problem’, success in the representation phase will lead to the creation of a quadratic equation and success in the solution phase is finding the factors of this equation. The core competency question associated with the Lawn Problem contained a (different) quadratic equation with instructions to find the factors. The set of core competency questions was the instrument to measure the problem solution phase. Hence, each participant was presented with the two tasks – a set of math problems first followed by a set of math questions. Since the problems were to be solved they required both representation and solution phases whereas the questions only required a solution phase. Problem representation thus became the subtraction of question from problem, i.e. if the participant had the math competencies to solve the problem, then success or lack of success in the problem was determined by the representation step. If the participant didn’t have the math competencies then failure to solve the problem could be due to failure in representation and/or failure in solution. Such cases could be removed before quantitatively examining the statistical relationship between spatial ability and problem solving while, for the qualitative analysis, all cases could be examined for type of representation regardless of success or failure in solution.

Following a post positivist epistemology, the hypothesis was recast as a null hypothesis: there will be no difference between weak and strong visualizers on measures of problem solving (i.e.

representation plus solution) and core competencies needed to solve the problems (i.e. problem solution), or rephrased as:

Strong visualizers will perform equally to weak visualizers on measures of problem solving and on measures of math competencies needed for the problems.

Data collection and analysis related to testing the null hypothesis are presented in this chapter. Presented first are the results of the pilot study that was used to develop a set of problems for the full study. While this revealed a large correlation between spatial ability and problem solving, the sample size was small and the pilot did not include a separate measurement of core competencies. However, it was instrumental in identifying the problems and questions used in the study. To test the null hypothesis, a set of math problems, a set of math questions and a spatial test were given to two samples of first year engineering students – one from DIT and one from OSU. These data were analysed to extract descriptive statistics and correlations between measures.

5.1 Developing and testing the problems and questions

Collating a set of math problems for this study was not as straightforward as first anticipated as the term 'problem' has a wide range of uses and connotations. Several sources of problems were reviewed, including: puzzle books, websites designed to teach mathematics or provide puzzles and problems, state examination papers from Ireland and the UK, a set of problems created at the University of Limerick called 'Inulis', academic books on thinking and problem solving and journal articles on the topic. The following criteria were used when selecting a problem:

- It was deemed to be a task that an engineering student should be able to tackle but was free of engineering discipline specific expertise so that it could be attempted by an engineering student from any discipline;

- It was not immediately obvious how to solve it;
- It required some analytical or critical thinking;
- It was more than just a question so that, when answered, a feeling of uncertainty would be aroused thereby prompting the participant to check the answer;
- The problem statement was brief and to the point and could be read in 20 seconds or so the problem could be solved in approximately 5 minutes so that a six problem test could be administered in approximately 30 minutes.

Problems in which equations were provided were not included as these could become tests of mathematical skills alone, rather than problem solving. A representation step is typically required in problems where only a description in words is provided, there is some puzzlement when reading the problem and the reader must think about what is meant and what to do before any test of mathematical skills occurs. This review of problems from the various sources led to the preparation of 15 problems for the pilot study which are found in Appendix A.

A convenience sampling approach was taken for this pilot study. The 13 students who volunteered to participate were all recruited from the teaching assistant pool of the Department of Engineering Education at OSU and were 2nd through 4th year students from various engineering disciplines. They were remunerated for their efforts with a gift card of small value. Ethical approval for the study was obtained in advance.

Participants were asked to attend an 'interview' that lasted 90 minutes divided up as follows:

- introduction to the project and completion of an informed consent form (10 mins)
- complete the MCT (20 mins)
- think aloud with LiveScribe while attempting the series of mathematics problems (60 mins)

A calculator was not allowed. Participants used a LiveScribe pen and paper which provides a concurrent recording of audio and written text. They were encouraged to think aloud while solving the problems and to articulate any decisions they were making. I conducted 9 interviews while a colleague conducted the remaining 4. We agreed on the protocol in advance and followed it for each interview. There was some variation in the selection and number of problems that each participant solved. Some of the problems (e.g. 'Cleaner' and 'Rectangle') emerged to be trivial for the participants and were dropped to be replaced by more challenging problems. There were 4 problems that were attempted by every participant with 5 more being attempted by at least 11 of the 13 participants. Table 5-1 presents the number of participants who attempted each problem along with success rates and solution times. There was variation among the participants in terms of which and how many problems were solved. For example, problems that were answered correctly by several participants in a row were deemed too easy and removed from further use (e.g. the 'Cleaner' problem). Problems were accepted if they split the sample somewhere in the middle and took around 5 minutes to complete.

Pilot study findings

The number of problems attempted by each participant ranged from 7 to 12. Expressed as a percentage of total number of problems attempted by each participant, the average score was 56 % with a range of 11 - 100 %. The large variation in success rates on the problems was a positive finding from the pilot study as it matched a desire to select problems that are discriminating. Likewise, the average MCT score was 15.1 (60.4 %) with a range of 6 to 24 (24 - 96 %). Scores on the mathematics problems were found to significantly correlate with the MCT, $r = .79$, $p < .01$, with 63 % of the variation in math scores shared with spatial ability. While this is regarded as a 'large' correlation, the sample size was very small in this case and even though the significance value indicates a low probability of this occurring by chance, there is no guarantee the same finding would emerge in a different sample.

Problem	Attempted by	Correct Answers	Incorrect Answers	% Correct	Average Time (mins)	Min Time (mins)	Max Time (mins)
Jug	13	3	10	23	5	1	11
Pizza	13	11	2	85	3	2	6
Cleaner	6	6	0	100	4	3	6
Sequence	12	9	3	75	3	0	6
Cans	13	9	4	69	5	2	10
Rectangle	2	2	0	100	3	2	4
Blood	11	8	3	73	4	2	7
Rain	12	6	6	50	7	3	15
Track	13	3	10	23	10	4	30
HCl	11	8	3	73	11	4	22
Fly	11	3	8	27	9	3	18
Egg	3	0	3	0	6	3	11
Lawn	4	2	2	50	9	3	14
Jars	4	3	1	75	4	2	10
Stadium	4	0	4	0	7	5	8

Table 5-1. Summary of performance data for problems used in pilot study of mathematical problems.

The think aloud protocol using the LiveScribe pen produced concurrently recorded spoken and written responses. The workbooks were examined for the use of sketching and the quality of sketches produced. The number of sketches produced was moderately correlated ($r = .42$, N.S.) with the score on the math test and more highly correlated ($r = .57$, $p < .05$) with score on the MCT although the quality of the sketches did not seem to follow any pattern with respect to spatial ability or math score.

Some minor, but important, issues related to the phrasing of the problem statements also emerged during the pilot phase. One participant was initially unsure of the meaning of a lawn so a definition was added in case others also lacked this semantic knowledge. The word “square” was bolded in the Lawn Problem statement as it was missed by some in their solutions.

Scale up to larger sample size

Based on the results from the pilot study, 6 problems were selected as most closely matching the criteria presented earlier, i.e. solution time was approximately 5 minutes, it appeared to be a problem to the students and success on the problem split the sample somewhere near the middle. The problems selected were Lawn, Jug, Cans, Rain, Pencils & Jars and Blood. A set of core competency questions was then developed to cover the math competencies needed to solve each of the problems. For example, the Lawn problem required the creation and solution of a quadratic equation. Hence, the associated core competency was identified - obtain the factors of a quadratic equation – and a question was created that presented a quadratic in x, with the same structure but different constants as the problem quadratic, and asked the participant to find the factors. This led to the creation of 6 core competency questions which can be found in Appendix B. The problem and question sets were now ready to be administered to larger samples of engineering students.

Sample 1 consisted of n = 62 first year engineering students at DIT, academic year 2015/16. In April 2016 they were administered the math problems and questions during a class period. Participants first read a consent form explaining the purpose of the project and the format of the test. They were asked to sign this test if they were willing to participate. They then worked individually on the problems and questions and I proctored or invigilated the test to ensure this happened. They were allowed 30 minutes to attempt all problems and questions and the use of a calculator was allowed. The consent form and math test were presented as a booklet in which they wrote their answer by hand. The consent form, provided in Appendix E, was on the front page followed by the problems and then the questions. Participants in this sample were members of a class group that had taken the Revised PSVT:R (Yoon, 2011) in the first semester of the same academic year, i.e. November 2015. While the MCT was used in the pilot study the PSVT:R was selected for the larger sample as it is administered during first year engineering orientation at both DIT and OSU. Hence, PSVT:R data were readily available and both MCT and PSVT:R have been shown to be strongly correlated with Farrell et al. (2015) reporting a correlation of $r(530) = .646$, $p < .01$, between the two tests with data collected from first year science and engineering students.

Problem	Sample 1 (DIT)	Sample 2 (OSU)
P1 Lawn	Written response	EEG & written response
P2 Jug	Written response	EEG & written response
P3 Cans	Written response	EEG & written response
P4 Rain	Written response	Audio think aloud & written response
P5 Pencils & Jars	Written response	Audio think aloud & written response
P6 Blood	Written response	Audio think aloud & written response
P7 Track	Not taken	Audio think aloud & written response
P8 HCl	Not taken	Audio think aloud & written response

Table 5-2. Problems administered and format of data collected from each sample

Sample 2 consisted of n = 54 1st year engineering students at OSU who were identified as 'weak visualizers' based on results of the PSVT:R (Guay, 1976) test administered during the 1st year orientation phase in Summer 2016. Between the middle of August and early September 2016, these participants were administered 8 math problems – the same 6 as the DIT sample plus 2 additional problems (see Table 5-2). A different protocol was used for these students as data were being collected for a related project at the same time. This study required that participants solve three problems while wearing an Electroencephalogram (EEG) hat. This still allowed sufficient data to be collected for this study, i.e. the additional data collection did not interfere with or compromise the data collection for this study. Each individual participant was invited to attend a one to one interview which consisted of two parts. In the first part, the purpose of the study was explained to each student and he/she was presented with an informed consent form to sign (see Appendix E). Then, the participant was asked to think aloud while solving 5 of the 8 problems and then answering the core competency questions. A LiveScribe pen and notepad was used to concurrently record audio and written response and 30 minutes was allowed for this part. The participant was then fitted with an EEG hat before commencing the second part in which he/she again used the LiveScribe pen but refrained from speaking so that an EEG recording could be made while the participant attempted the next 3 problems. Fifteen minutes was allowed for completing the 3 problems. One OSU participant consented to participate only in the think aloud session. It was not possible to conduct the

EEG session with this participant on the same day and the participant failed to return. Hence, for 3 of the problems (1, 2 and 3) there were 53 and not 54 OSU participants.

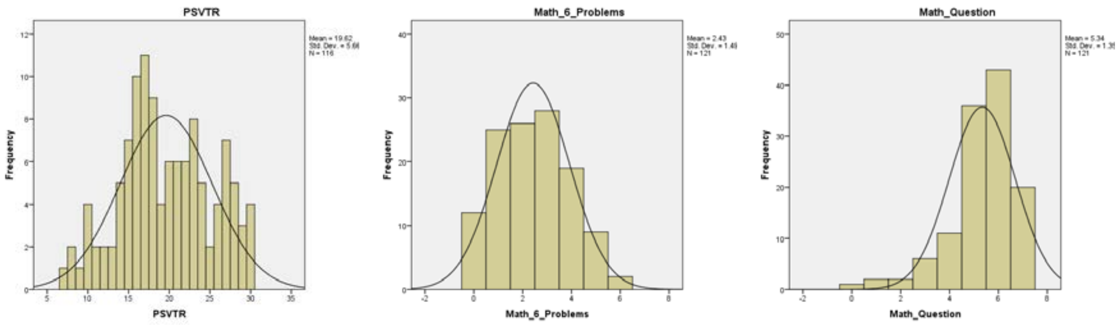
During the think aloud session, I and the participant sat at a small table and I presented one problem at a time. I asked the participant to think aloud while working and to try to articulate decisions and dilemmas rather than a running commentary of what he/she was writing. I only reminded the participant to think aloud if he/she remained quiet for longer than 30 seconds. I also kept an eye on the time. If the participant spent more than 5 minutes on a problem and was close to completing it I refrained from speaking. If the time limit was exceeded and no solution was imminent I gently suggested he/she wrap up and move on to the next one. Participants were not allowed to use their own calculators in the belief that this would encourage more independent thinking. If a participant was struggling with arithmetic I intervened and offered the use of my calculator. Another researcher conducted the EEG recording sessions.

The PSVT:R tests were scored using the answer key provided by the test authors. The math questions and problems were scored according to the solution sheets. If a solution contained the correct numerical answer it was scored as correct. If it contained an unfinished computation that would have provided the correct numerical answer were the computation finished then it was also scored as correct. In other words, the use of a calculator would have been sufficient to make the last step in the solution. For example, for the Blood problem, all of the following answers were scored as correct: 42.4 cm^3, 42 cm^3, 42, 13.5π, πx1.52x6. The only difference between these answers is the extent of arithmetic computation.

Statistical properties of the instruments

Some statistical properties of the frequency distribution of the data were measured in order to check for normality of the data from the PSVT:R, the math problems and the math questions. These results are presented both graphically (Figure 5-2) and numerically (Table 5-3). Both the

PSVT:R and the math problem data contain some skewness and kurtosis but not to a significant extent (when divided by their respective standard errors the values are less than 1.96 in all cases (Field, 2013)). The math questions data, however, contain significant amounts of skewness and kurtosis. The PSVT:R and math problem data can be considered to be normally distributed whereas the math question data cannot.



(a) (b) (c)

Figure 5-2. Frequency plot of data collected from DIT and OSU 1st year engineering samples for (a) PSVT:R, (b) math problem and (c) math questions.

	PSVTR	Math Problems	Math Question
N	115	115	115
Mean	19.70	2.42	5.31
Std. Deviation	5.613	1.481	1.372
Minimum	7	0	0
Maximum	30	6	7
Kurtosis	-.671	-.629	2.385
Std. Error of Kurtosis	.447	.447	.447
Skewness	-.004	.222	-1.309
Std. Error of Skewness	.226	.226	.226
$Z_{Kurtosis}$[1]	-1.501	-1.407	5.336*
$Z_{Skewness}$[1]	-0.018	0.982	-5.792*

[1] $Z_{Kurtosis}$ = Kurtosis/Std. Error, likewise for skewness
* Significant, hence, not normally distributed

Table 5-3. Statistical properties of data collected from DIT and OSU 1st year engineering samples for PSVT:R, math problem and math questions.

Descriptive statistics of the two samples

With regard to the two samples, some descriptive statistics are presented separately for each sample in Table 5-4 and for each of the measures – PSVT:R, math problems and math questions – along with the numbers of weak (PSVT:R ≤ 18) and strong visualizers (PSVT:R ≥ 19).

Sample	n	PSVT:R Mean	S.D.	Math Problem Mean	S.D.	Math Question Mean	S.D.	n Weak	n Strong
DIT	62	22.37	5.53	2.85	1.48	5.00	1.22	10	52
OSU	53	16.58	3.86	1.91	1.32	5.68	1.47	45	8

Table 5-4. Descriptive statistics for the two samples

As can be seen in Table 5-4, the OSU sample has a much lower average spatial test score because it predominantly included those who had 'failed' the PSVT:R taken during orientation. No such restriction was imposed when recruiting participants to the DIT sample. It contained 10 weak visualizers, or 16 % of the sample, which is typical for samples of 1st year engineering students. A similar proportion (17 %) of weak visualizers was measured using the full set of data (n = 1053) collected during the orientation at OSU in summer 2016 (Duffy, Sorby, Mack, & Bowe, 2017). From this point on the two samples were combined and treated as one.

5.2 Summary of results for problems and core competency questions

To begin addressing the hypothesis that problem representation and not problem solution is connected to spatial ability, Pearson correlation coefficients between scores on all three variables were calculated. The sample was then grouped by spatial ability into weak and strong visualizers to compare the problem and question scores of this dichotomous grouping. This analysis of math scores was followed by looking at performance on the individual questions and problems. Two approaches to analysing math performance were taken. One was to group the sample into correct and incorrect groups and compare the PSVT:R scores of each group, calculating the magnitude and significance levels of any differences. The second was to calculate point-biserial correlation coefficients between the score on the problem or question (either 0 or 1) and the PSVT:R score.

Correlations between measures

A strong and highly significant correlation coefficient was measured between the PSVT:R and the math problem scores: $F(1, 114) = 47.42$, $r(113) = .544$, $r^2 = .296$, $p < .001$ (Table 5-5). In contrast, there was a negligible and insignificant correlation coefficient measured between the PSVT:R and the math questions: $F(1, 114) = .1.99$, $r(113) = .131$, $r^2 = .017$, N.S. Lastly, the math problem scores were found to have a significant correlation with the questions: $F(1, 114) = 23.974$, $r(113) = .418$, $r^2 = .175$, $p < .001$. The Pearson correlation is based on the assumption that the data are normally distributed which is violated in the case of the math question data. It is more appropriate, therefore, to use the non-parametric Spearman correlation coefficient and this was calculated as $r_S = .459$, $p < .001$ for math problems to questions and as $r_S = -.153$, N.S. for the PSVT:R to questions. With either method – Pearson or Spearman - the correlation coefficient between the two math measures is significant with 18 to 21 % of variation shared between them.

	Math Problems **r**	**Math Questions** **r**	**Math Questions** **r_S (Spearman)**
PSVT:R	.544***	.131	.153
Math Problems		.418***	.459***

** significant at $p < .01$
*** significant at $p < .001$

Table 5-5. Correlation matrix for scores on the PSVT:R, math problems and math questions, n = 115.

These findings indicate a stark contrast in magnitude and significance between the PSVT:R correlations with each math measure: 30 % of the variation in the PSVT:R data is shared with the problem scores but nothing of significance is shared with the question scores. This difference indicates that the null hypothesis – there is no difference between weak and strong visualizers on either math measure – is false. The correlation values tell us that as PSVT:R scores increase, math problem scores also increase but math question scores remain constant.

Hence a grouping based on two different ranges of PSVT:R should reveal a difference in math problem scores but not question scores. This was tested by comparing the math problems scores of weak and strong visualizers using an independent samples t-test which revealed a sizeable and significant difference between the math problem scores of the two groups but not between their math question scores (Table 5-6). Hence, the null hypothesis can be rejected and it can be said that there is a difference between weak and strong visualizers in problem solving ability but not in core competency skills.

	Weak			Strong					
Math Measure	**n**	**M**	**SD**	**n**	**M**	**SD**	**t-test**	**Sig (2-tailed)**	**Cohen's d (Size)**
Problems	55	1.71	1.36	60	3.07	1.29	-5.507	.000	1.03 (Large)
Questions	55	5.49	1.48	60	5.15	1.26	1.335	.185	0.25 (Medium)

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Table 5-6. Sample grouped by spatial ability level to compare means of math problems and questions.

Given the relative simplicity of these problems, this finding can be interpreted to mean that spatial ability plays an important role during problem representation but is not relevant for the problem solution phase.

Performance on each problem and question

Some further statistical analysis was conducted to examine how weak and strong visualizers performed on each individual problem and question. This was done by grouping the entire sample by correct or incorrect response to each problem and question and then calculating mean PSVT:R scores for each of these two groups as shown in Table 5-7 for the problems and Table 5-8 for the questions. These means were then compared using an independent samples t-test to measure the significance of any difference between them. A Cohen's d effect size was also calculated to indicate the relative size of the difference.

	Correct			Incorrect					
Problem/ Question	n	M	SD	n	M	SD	t-test	Sig (2-tailed)	Cohen's d (Size)
P1 Lawn	30	21.53	5.53	85	19.06	5.53	-2.107	.037*	0.45 (Medium)
P2 Jug	69	21.59	4.93	46	16.87	5.43	-4.837	.000**	0.92 (Large)
P3 Cans	45	22.16	4.73	70	18.13	5.60	-3.993	.000**	0.78 (Large)
P4 Rain	33	22.67	4.35	80	18.44	5.79	-3.776	.000**	0.83 (Large)
P5 Jars	43	19.19	5.24	72	20.01	5.84	.764	.447	0.15 (Small)
P6 Blood	62	21.76	4.71	53	17.30	5.67	-4.604	.000**	0.86 (Large)
P7 Track	3	17.00	1.00	46	16.41	4.20	-.239	.812	0.2 (Small)
P8 HCl	9	20.67	4.87	20	14.45	2.80	-3.571	.005**	1.57 (Large)

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

Table 5-7. Comparison of means of PSVT:R scores for both the DIT common 1st eng and OSU freshman eng groups for those correct and incorrect on each math problem (n=115)

For problems 1 to 6 (Table 5-7), the differences between the spatial test scores of the Correct and Incorrect groups are significant at the $p < .01$ level for 4 problems – Jug, Cans, Rain and Blood,significant at the $p < .05$ level for the Lawn Problem and not significant at all for the Jars problem. In terms of effect size, the magnitudes of the differences in PSVT:R was found to be large for all but the Jars problem. For Problems 1 to 6 there are at least 30 participants in each group (correct v incorrect). Problems 7 and 8 consist of fewer cases as these problems were not attempted by the DIT sample. It was decided that the membership of these categories is too small to draw meaningful conclusions from statistical results and they were excluded from further analysis in this report.

	Correct			Incorrect					
Problem/ Question	n	M	SD	n	M	SD	t-test	Sig (2-tailed)	Cohen's d (Size)
Q1 (P1)	89	20.34	5.551	24	17.21	5.80	-2.427	.017*	0.56 (Large)
Q2 (P2, 4, 6)	86	20.48	5.47	29	17.41	5.49	-2.605	.010*	0.57 (Large)
Q3 (P3, 4)	108	19.81	5.71	5	16.80	5.89	-1.149	.253	0.52 (Large)
Q4 (Q2)	95	19.65	5.53	18	19.78	6.83	.085	.933	0.03 (Small)
Q5 (P4)	90	19.10	5.37	23	21.91	6.61	2.136	.035*	0.47 (Medium)
Q6 (P5, 8)	95	19.76	5.703	18	19.22	5.99	-.363	.718	0.1 (Small)
Q7 (P7)	40	16.43	3.46	12	17.17	5.24	.575	.568	0.17 (Small)

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

Table 5-8. Comparison of means of PSVT:R scores for both the DIT common 1st eng and OSU freshman eng groups for those correct and incorrect on each math question (n=115)

In contrast, when the same comparison was made between the Correct and Incorrect groups for the core competency questions (Table 5-8), none of the differences in PSVT:R were significant at $p < .01$, three were significant at $p < .05$ – questions 1, 2 and 5 – and three were not significant at all. As measured by Cohen's d, moderate to large and significant differences were observed for 2 of these questions. It was interesting to observe some significant differences in spatial ability based on these simple tests of math competencies as there was no significant correlation between questions as a set and spatial ability. This raised the possibility that some participants failed to solve the problems because they lacked the required core competency. To eliminate core competency as potentially confounding variable, the sample was restricted to those who correctly answered the corresponding core competency question when analysing the data for each problem. This repeated analysis is presented in Table 5-9 where it can be seen that the reduction in the number of cases was not significant because a majority of participants correctly answered the core competency questions.

	Correct			**Incorrect**					
Problem	**n**	**M**	**SD**	**n**	**M**	**SD**	**t-test**	**Sig (2-tailed)**	**Cohen's d (Size)**
P1 Lawn (Q1=1)	27	22.30	5.74	62	19.48	5.29	-2.247	.027*	0.52 (Large)
P2 Jug (Q2=1)	64	21.88	4.85	22	16.41	5.24	-4.470	.000**	1.09 (Large)
P3 Cans (No Q)	44	22.27	4.72	71	18.11	5.56	-4.125	.000**	0.81 (Large)
P4 Rain (Q2 & 5=1)	27	22.41	4.45	41	18.27	5.48	-3.277	.002**	0.83 (Large)
P5 Jars (Q6=1)	36	19.03	5.33	60	20.30	5.73	1.080	.283	0.23 (Small)
P6 Blood (Q2=1)	59	21.81	4.80	27	17.30	5.97	-3.744	.000**	0.84 (Large)

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

Table 5-9. Comparison of means of PSVT:R scores for those correct and incorrect on each math problem with cases excluded if the answer to the question corresponding to the problem is incorrect (n varies by problem).

After limiting the sample in this way, the pattern observed in Table 5-7 is repeated in Table 5-9. Membership of each group is still reasonably large, the smallest containing 22 participants. As before, there is a highly significant ($p < .01$) difference with a large effect size ($d \geq .75$) in the spatial ability levels between correct and incorrect problem solvers on the Jug, Cans, Rain and Blood problems, a significant ($p < .05$) and moderate ($d = .52$) difference on the

Lawn Problem and no difference on the Jars Problem. Again, the Jars and Pencils problem (no. 5) is the exception to the rule – there is no difference in spatial ability levels of those who got this problem correct and those who didn't.

In a case like this where one variable is binary or dichotomous and the other continuous, an alternative is to calculate a point bi-serial correlation by grouping the sample as above and comparing the PSVT:R scores of the two groups. These results, presented in Table 5-10, support the same findings as above: correlations are medium to large and significant at $p < .01$ for Problems 2, 3, 4 and 6, small for the Lawn problem while the Pencils & Jars problem is the exception with no significant correlation.

Problem	1	2	3	4	5	6
PSVT:R	.203*	.414**	.362**	.336**	-.072	.395**

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

Table 5-10. Point bi-serial correlation between PSVT:R and each math problem (n = 115)

The correlation between the spatial test and math problem scores can also be presented in a categorical way by grouping the sample based on the number of correct problems scored and then computing the mean PSVT:R score for each of these groups. As shown in Table 5-11 and Figure 5-3, weak visualizers are overrepresented in the low scoring groups and underrepresented in the high scoring groups.

				PSVT:R	
Score on Problems	**Weak**	**Strong**	**Total**	**Mean**	**S.D.**
0	11	0	11	11.73	3.636
1	17	7	24	17.17	3.818
2	12	14	26	20.15	5.167
3	9	17	26	21.65	5.628
4	4	14	18	22.33	3.614
5	2	6	8	22.50	4.660
6	0	2	2	28.00	2.828

Table 5-11. Sample grouped by success rate on problems and by spatial ability.

The problems the participants were asked to solve are short, simple algebra word problems that require basic arithmetic and either a basic core math competency or prior knowledge of an equation such as cylinder volume. Problems of this nature are typically found in a US middle school or the Irish junior certificate curriculum. In order to solve the problem, a participant is required to read and interpret the problem statement from which a strategy or approach is determined before applying basic arithmetic, core competency and/or prior knowledge. The findings from this analysis show the entire problem solving process shares much in common with the ability to score well on a totally different test, the PSVT:R, that measures a cognitive factor called spatial visualization.

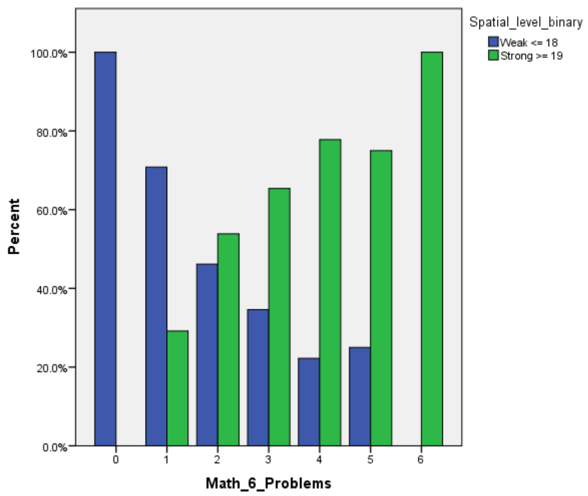


Figure 5-3. Number of weak and strong visualizers grouped by score on the math problems and expressed as percentage of total for each problem score.

The absence of a correlation between the PSVT:R and the competency questions can be interpreted in two ways – there is no relationship between the two measures or there is a relationship but it is hidden by the homogeneity of the question data set. No relationship implies the cognitive activities required by each group are very different and do not overlap. This is plausible if these competencies have been practiced so many times that they can be applied in a reflexive manner with little thought so one is not required to visualize or hold things in working memory while transforming them. The lack of variation in the competency data set itself (most participants were correct on most of the questions) could result in no correlation when there potentially is a relationship but it's just not revealed by these particular competency questions. Despite this homogeneity, a significant but small correlation did emerge with the problem solving scores. The two math measures are certainly related – one can't solve the problem without having the core competency – and are strongly enough related for a correlation to be observed even with one data set being highly skewed. Therefore, if a relationship exists between competency questions and PSVT:R, it is not strong enough to emerge in this case and it seems more logical to conclude the two measures share little in common at a cognitive level. As discussed in the literature review, not all measures of mathematics correlate with spatial ability.

These findings reveal a correlation between spatial ability and problem solving and a significant difference in the problem solving abilities of weak and strong visualizers. This difference is not explained by the problem solution step as there is no variation in core competency with spatial ability. When cases are limited to those who correctly answered the competency questions, the findings still hold. Therefore, the null hypothesis is falsified on the basis of these results – it has been shown there is a difference between weak and strong visualizers in the way they solve problems. If the null hypothesis is false, does this mean the hypothesis is true, that strong visualizers have superior skills in representing problems? Not necessarily. In fact, all that has been shown is that when this sample is grouped by spatial ability scores a difference was observed in the math problem scores of the two groups. Why there is a difference remains a matter of conjecture and is the point addressed in the next round of analysis, a qualitative examination of the participants' solutions to identify how participants were reading and interpreting the problems, what actions they performed in solving them, what mistakes they made and what schema were evident in their solution.

Repeatability of the findings

It is quite possible these findings may not hold for other samples of first year engineering students. This could occur, for example, if the problem solving data collected from another sample were highly skewed, either negatively (sample consists of very good problem solvers) or positively (bad problem solvers). In such a case, a correlation with the spatial test might not emerge and very different conclusions would be made. This occurred in one multi sample study in which a significant relationship between spatial and math was observed for some samples but not for others, i.e. it varied with the sample (Casey et al., 1995). Hence, there is a real risk that these findings might not be observed in other samples and that what has been observed here is unlikely to be repeatable in other contexts.

In order to address the issue of repeatability, findings from an additional two samples are presented here and compared with the existing results. However, some interpretation is required as there was variation in the measurement instruments used. An additional sample from DIT consisted of 23 second year engineering technology students who were administered the MCT and the 6 math problems and 6 math questions. Another sample from OSU consisted of the majority of first year engineering students who completed the PSVT:R and a Math Placement Test (MPT) during their orientation programme in the summer of 2016. The MPT consists of 25 questions, many of which are procedural in nature but several questions require some form of problem solving. One question on the test was in word algebra format. An example of this test is provided in Appendix D. Of the 25 questions, five were categorised as problems and 6 as core competency questions. The remaining 14 did not fit into either category. Correlations between PSVT:R and all 25 questions on the MPT, the five problems and the 6 questions were then calculated. The findings from these additional samples are shown along with the existing data in Table 5-12.

Sample	Spatial Test (Mean, S.D.)	Date	n	r Overall	r Problems	r Questions
OSU engineering	MCT (15.08, 5.97)	January 2016	13	N/A	.790**	N/A
DIT 2nd Year B. Eng. Tech.	MCT (9.22, 4.37)	April 2016	23	.574**	.592**	.393
DIT 1st year common engineering	Redrawn PSVT:R (22.37, 5.53)	April 2016	62	.445**	.525**	.189
OSU Freshman Eng Low Spatial	PSVT:R (16.39, 4.02)	August 2016	53	.270	.332*[1]	.024
OSU Freshman Eng All students	PSVT:R (23.58, 4.51)	August 2016	1053	.207**	.233**	.088**

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).
[1]. An additional two math problems were administered to this sample but this correlation was calculated using data from the set of 6 problems that were common to both samples. The correlation between the PSVT:R and all 8 problems was measured to be .387**, i.e. a very similar result.

Table 5-12. Summary of spatial-math correlations for all samples used in this study.

One interesting finding did emerge from the MPT data. When each question was analysed by comparing the PSVT:R scores of those who were correct and incorrect, the largest effect size belonged to the word algebra problem on the MPT ($d = .41$, $p < .000$) and effect sizes were found to be negligible for the procedural questions.

The finding taken from the DIT and OSU 1st year engineering data – spatial correlates with problem solving but not core competencies – was replicated using data from another DIT sample and a very similar observation was made using data from a large sample of OSU 1st year engineering students which showed a sizeable and significant difference in spatial ability for those able to solve a word algebra problem versus those who could not. While repeatability is never guaranteed, some evidence exists to show these findings are not isolated but are likely to be observed in other samples.

Conclusions

Beginning with a broad and somewhat confused definition of engineering problem solving, the initial phase of this project was successful in producing a measurement of problem solving that co-varied to a large and significant extent with the measures of spatial ability selected for the study. This success was both notable and encouraging as not all measures of mathematical ability correlate with spatial ability tests. The set of math problems produced in the pilot study presented an opportunity to collect valuable data that could be used to examine the relationship between math problem solving and spatial visualization. Given the word algebra nature of the problems it also presented an opportunity to add to the existing literature on approaches to solving such problems.

The two samples of students who participated in this study came from first year engineering programmes in different locations, Dublin Institute of Technology and Ohio State University. Although the percentage of weak visualizers, those scoring 18 or lower on the PSVT:R, in such samples is normally around 20 %, the focused recruitment of weak visualizers in the OSU sample resulted in a combined sample that contained 55 weak and 60 strong visualizers. This was important from a statistical point of view because when the sample was divided in various ways during the analysis the membership of the resulting groups was, in most cases, large enough to allow statistical tests of significance to be performed.

Having administered the PSVT:R, the set of math problems and the core competency questions to the full sample ($n = 115$), it was found that both the PSVT:R and the math problems scores were normally distributed. Most participants achieved high scores in the set of competency questions which led to this data set being highly skewed to the upper end of the scale. The PSVT:R was found to be significantly correlated with scores on the math problems, $r = .544$, $p < .001$. The magnitude of this correlation indicates a sizeable amount of variation, 30%, is shared between the two measures or, in other words, 30 % of the variation in the math scores is explained by the spatial measure. A significance of $p < .001$ indicates it was highly unlikely this occurred by chance and would be observed again if the study was repeated with this sample. Scores on the core competency questions were found to be significantly correlated with the math problems, $r = .289$, $p < .01$, but the magnitude of this correlation is small with only 8 % of variation shared. Math competencies are needed to solve the problems but the core competency data set is too skewed for a large correlation to emerge. No significant correlation between the math questions and the PSVT:R was observed and it is concluded that these two measures share little in common at a cognitive level. These finding were replicated in another sample of engineering technology students from DIT.

Analysis of a large Math Placement Test data set collected from 1053 1st year engineering students at OSU revealed that correlations with the PSVT:R were very small for questions that were highly procedural and larger for questions that had an element of problem solving. The largest correlation was found for the one question on the MPT that was presented as a word algebra problem. Hence, it is concluded that the results presented in this study are likely to be replicated if the same measures are administered to other samples of engineering students in other contexts and settings.

Strong visualizers outperformed weak visualizers in solving all but one of the math problems, the exception being the Pencils and Jars problem. When grouped by correct and incorrect for each problem, large and highly significant differences in spatial ability were found for the other five problems. When these comparisons were repeated for each problem after excluding those who did not answer the corresponding core competency question correctly the same results were found, meaning that lack of success in problem solving was not explained by a lack of core competency. Strong visualizers were more successful at problem solving for reasons other than core competency in mathematics. The null hypothesis was falsified – there *are* differences between weak and strong visualizers in problem solving despite there being no differences in core competency.

In summary, this analysis identified a strong relationship between spatial ability and problem solving that could not be explained by mathematical ability but was connected with the way problems were being represented by the participants. A second round of analysis was now required in order to understand participants' approaches to solving each problem, what representations they were creating at the linguistic, semantic and schematic levels and the steps they were taking to solve the problems. An objective of the next stage of analysis was to identify the errors made by participants when representing and solving the problems and to rank these errors by spatial ability in order to reveal why strong visualizers outperformed weak visualizers on five of the problems but not on one problem.

Chapter 6 Spatial ability, problem representation and problem solution

Introduction

The aim of the second round of data analysis was to learn about the qualitative variation in the responses from the 115 DIT and OSU participants as part of the larger goal of trying to explain why visualization ability predicted success in problem solving. This initial round of qualitative data interpretation was restricted to the written data and did not include the audio recordings. Since written responses were obtained from both DIT and OSU participants the sample size in this phase was also n = 115. By examining these responses the intention was to find reasons why some problem solvers were successful and others were not. The approach taken was to interpret participants' solutions by identifying problem solving steps as enacted in their solutions, i.e. the actions taken in solving the problems. Once a full list of actions was identified, the presence or absence of each potential action could be counted for each participant and then summed for all participants. This would then allow the grouping of the sample by successful and unsuccessful actions or approaches and a statistical comparison of spatial ability between these groups. Having left the previous analysis only knowing how many were correct and incorrect on each problem, this next round of analysis aimed to provide some description of the paths towards correct and incorrect solutions.

6.1 Coding scheme development illustrated using the Lawn Problem

Two alternative viewpoints were employed when interpreting the participants' solutions to the Lawn and the other problems – one was to adopt an established description of problem solving presented by Mayer (1992) and the other was to start with a blank slate and be open to whatever was contained in the data. The latter is generally more challenging as not knowing where to start makes it harder to start. The attraction of Mayer's schema was that it existed and was developed for simple word algebra problems and could be applied to the problems used in this study to develop a set of a priori codes which could be used as data were examined. The weakness of Mayer's schema was that it does not allow for variation in problem solving approaches and representations that may be naturally present in the data. As the data were examined through the lens of the Mayer framework, actions that might not be contained in or consistent with the a priori Mayer schema were also searched for and noted.

Analysis was completed and is presented separately for each problem beginning with the Lawn problem, for which the analytical process is elaborated to a greater extent in order to illustrate the process in detail. The order in which the problem analyses are presented in this chapter and the next is based on the order in which they were presented to the students. The Lawn problem was the first to be analysed and, therefore, was accompanied by a larger learning curve but this ordering of the problems is not important as each problem was analysed separately and independently. With the exception of the Lawn problem as it used to elaborate the method of analysis these analyses can be read in a different order to which they are presented if the reader so desires. Since this chapter and the next are concerned with the analysis of the problem representation phase data collected from all 115 participants were included as it was assumed that all were capable of attempting to represent the problem even if they could or could not correctly answer the core competency questions. In other words, a lack of ability to complete the solution step was not assumed to hinder an attempt to complete the preceding representation step. As mentioned in Chapter 5, participants were presented with the full set of problems first and then then the full set of questions.

Step 1, Apply Mayer's schema to develop the first set of codes

The Lawn problem is presented first to provide some context for the coding schemes:

> *"A **square** lawn was extended in width by 2 m and in length by 3 m. The area of the new lawn is twice as big as the area of the old lawn. What are the measurements of the old lawn?"*

The core competency question related to this problem is Question 1, i.e., factorise a quadratic equation:

Find the roots of $2x^2 + 6x - 8 = 0$ using factoring

The Mayer framework (Mayer, 1992) for math problem solving consists of several types of knowledge that are required to solve math problems. In the context of the Lawn problem, the application of each of these types of knowledge by the successful problem solver is outlined in Table 6-1.

Type of knowledge	Application to the Lawn problem
Linguistic knowledge	ability to understand the words used in the problem, e.g. the lawn is square, area of new lawn is twice the old area, new width equals old width plus 2, new length equals old length plus 3
Semantic knowledge	ability to draw on common sense or knowledge that is taken for granted, e.g. a square has four equal sides, a lawn is an area of grass beside a house (this definition was provided in the problem statement) that is 2 dimensional (not provided)
Schematic knowledge	ability to draw on knowledge of schema, i.e. that have been previously learnt; in this case, a schema that is required is that area = length x width
Strategic knowledge	ability to set subgoals in the problem solving process and to monitor progress, in this case one should develop an equation for one unknown – the side of the original square – and then solve for this unknown using algebra
Procedural knowledge	ability to perform standard mathematical procedures; in this case, an important procedure is to factorise a quadratic, (hence the core competency question testing this skill)

Table 6-1. Mayer framework (Mayer, 1992) for math problem solving applied to the Lawn problem.

Several problem solving actions can be performed when drawing on each of these types of knowledge. These actions were listed separately to allow for cases where a participant demonstrated some but not all actions that draw on linguistic knowledge, for example. This led to the creation of a more fine grained set of codes presented in Table 6-2.

No.	Code	Present if…. (1)	Absent if …. (0)	Type of knowledge	Required to solve problem
1	Square Lawn	Discerns lawn is a square	Does not, e.g. rectangle	Linguistic Assignment	Yes
2	Area change	Discerns A_{New}=2xA_{Old}	Does not	Linguistic Relational	Yes
3	Width+2	Discerns new width = old + 2	Does not	Linguistic Relational	Yes
4	Length+3	Discerns new length = old + 3	Does not	Linguistic Relational	Yes
5	2D object	Lawn is a 2D object	Does not	Semantic	Yes
6	A=WxL	Includes Area = width x length	Does not	Schematic	Yes
7	Correct equation	The correct equation is provided	Not provided	Strategic	Yes
8	Solve	Solve equation correctly	Makes error	Procedural	Yes
9	Check	Checks answer	Does not	Strategic	No

Table 6-2. A priori coding scheme for the Lawn Problem based on the Mayer schema.

Step 2, Start reading, coding and reflecting on codes

Each transcript was studied in turn checking for evidence of these actions in each participant's written solutions. If, on moving to the next solution script, a new action that was qualitatively different to an existing code was identified it was added as a new code to the list. These codes were phrased almost entirely as binary statements or questions, i.e. present or not present in the solutions. If there was more than one component to the code then it was split up to create several related binary codes or, in rare cases, a code was allowed to have several categories. Hence, the a priori list of codes changed with reflection and careful consideration while reading the solutions in a constant effort to replace, improve or discard codes. When changes occurred they were retrospectively applied to all participants' solution by returning to the first participant's solutions and scoring it and each subsequent participant for the modified codes.

In the case of the Lawn problem this process led to the creation of additional codes (Table 6-3) and the deletion of some a priori codes if they appeared to be redundant. For example, since every solution showed evidence of Code 3, 'Width + 2', also contained Code 4, 'Length + 3', it was possible to delete Code 4 without losing any information. All treated the lawn as a 2 dimensional object so Code 5 was also deleted. Finally, some codes (no. 15 to 18, Table 6-3) were created by combining other codes that connected together to form a coherent approach to solving the problem such as having all the ingredients to write the correct quadratic equation.

It was difficult to discern strategic knowledge from the answer sheets, particularly the monitoring aspect. Some solutions were very straightforward, simple and error free – there must be monitoring in action here but how could it be discerned? Other solutions started with a rectangular lawn and were then corrected to a square lawn – is this an example of monitoring sparked by an 'I can't solve this, I've done something wrong, let me go back' moment or might the student have just glanced again at the problem and the word 'square' caught the eye? Analysis of the data focused on the problem representation step, the topic of interest in this study, rather than this more difficult to interpret aspect of strategic knowledge.

No.	Code	Present if…. (1)	Absent if …. (0)	Type of knowledge	Required to solve problem
10	Schema change	Changes from incorrect to correct	No change	Strategic (is monitoring work)	No
11	L=W=x	Assigns a variable ID like x to length/width	Does not	Strategic	Yes
12	Number of equations	Has 1 equation for each unknown	Does not, e.g. 1 equation for 2 unknowns	Procedural	Yes
13	Guess & check	Uses guess and check approach	Does not	Strategic?	No
14	Equation equality	Both sides of equation are equal	Both sides are not equal	Procedural	Yes
15	Linguistic knowledge 1	Addition of Codes 1 & 3	Scored as 0, 1 or 2	Linguistic	Yes
16	Linguistic knowledge 2	Addition of Codes 1, 2 & 3	Scored as 0, 1, 2 or 3	Linguistic	Yes
17	Equation ingredients	Addition of Codes 1, 2, 3 & 6	Scored as 0, 1, 2, 3 or 4	Linguistic and schematic	Yes

Table 6-3. A posteriori codes for the Lawn Problem.

Statistical analysis of codes

Since the purpose of this phase of the analysis was to examine variation in approach to problem solving among weak and strong visualizers, the extent to which each of the codes revealed a difference in spatial ability was checked in two ways. First, the sample was grouped by spatial ability into weak and strong visualizers before counting how many of each scored 0 and 1 on each code. For example, the number of weak and strong visualizers who showed evidence of treating the lawn as having a square shape (code 1 = 1) was noted. Second, the sample was regrouped based on the code being present or not (0 or 1) and the mean and S.D. of the PSVT:R scores were computed for each group and compared using an independent samples t-test and a Cohen's d effect size. For example, all those who showed evidence of treating the lawn as a square were placed in one group (code 1 = 1) with the remainder (code 1 = 0) in the second group, the mean PSVT:R scores of each group were then compared using an independent samples t-test to measure the significance of any difference and a Cohen's d effect size was calculated to indicate the magnitude of the difference. This process was repeated for all the codes and the results are presented in Table 6-4.

The lawn problem contains several statements that define variables or relationships between variables which are contained in codes 1, 2, 3 and 6 and which varied greatly in effect size. While code 1, 'the lawn is square', divided the sample into two large groups, it resulted in a small to medium effect size ($d = .28$, N.S.) in terms of spatial ability difference between those who did and didn't treat the lawn as a square when solving the problem. Larger and more significant effect sizes were observed for codes 2 and 3, the relational statements in the problem ($d = .64$, $p < .05$ and $d = .72$, $p < .01$). Code 6, Area = W x L, produced the largest effect size ($d = .93$, $p < .05$) and all but eight participants showed evidence of this code in their solutions. These four codes were needed as a complete set to write the correct quadratic equation, i.e. it was essential to have them as a set to solve the problem. This combination is shown by code 17a which revealed a medium effect size ($d= .54$, $p < .01$).

No.	Code	=1 if	Weak n 1	Weak n 0	Strong n 1	Strong n 0	Total n 1	Total n 0	PSVT:R 1	PSVT:R 0	t (p)	Cohen's d
1	Square lawn	Lawn is square	31	24	42	18	73	42	20.27 (5.38)	18.71 (5.93)	-1.441 (.152)	0.28 (Medium)
2	Area change	$A_{new}=2xA_{old}$	43	12	51	9	94	21	20.34 (5.49)	16.86 (5.41)	-2.637* (.010)	0.64 (Large)
3	Size change	W+2, L+3	41	14	53	7	94	21	20.43 (5.36)	16.48 (5.72)	-3.017** (.003)	0.72 (Large)
6	Area	A=WxL	49	6	58	2	107	8	20.05 (5.52)	15.13 (5.06)	-2.444* (.016)	0.93 (Large)
7	Correct Equation	Has correct equation	9	46	22	38	31	84	22.13 (5.26)	18.81 (5.50)	-2.904** (.004)	0.62 (Large)
8	Solve	Solves equation correctly	6	49	17	43	23	92	22.43 (5.52)	19.02 (5.46)	-2.678** (.009)	0.63 (Large)
12	Number of equations	1 eqn. per unknown	17	38	29	31	46	69	20.96 (5.56)	18.87 (5.53)	-1.978 (.050)	0.38 (Medium)
13	Guess & Check	Uses guess & check	22	33	17	43	39	76	18.41 (5.24)	20.37 (5.72)	1.788 (.076)	0.36 (Medium)
14	Equation equality	Eqn. has equality	24	31	44	16	68	47	21.12 (5.77)	17.66 (4.74)	-3.394** (.001)	0.66 (Large)
15	Linguistic knowledge 1	Codes 1 & 2	25	30	38	22	63	52	20.87 (5.12)	18.29 (5.90)	-2.514* (.013)	0.47 (Medium)
16a[1]	Linguistic Knowledge 2	All 3 present	21	34	33	27	54	61	21.22 (5.24)	18.36 (5.63)	-2.810** (.006)	0.53 (Large)
17a[1]	Equation ingredients	All 4 present	20	35	32	28	52	63	21.29 (5.29)	18.40 (5.58)	-2.833** (.005)	0.54 (Large)

* Significant at the $p < .05$ level
** Significant at the $p < .01$ level
[1] Binary version of original code, must have all components to score as 1, e.g. if Code 16 = 3, then Code 16a = 1, else Code 16a = 0

Table 6-4. Statistical data for the codes in the Lawn Problem.

While 52 participants obtained the four ingredients required to write the equation, only 31 were successful in writing the correct equation and, of these, only 23 were successful in the solving the equation. At each of these steps, particularly the first two, strong visualizers were more successful than weak visualizers as shown in Table 6-5. The gap between weak and strong in solving the equation is not that large – if provided with the equation, 6 of 9 weak visualizers can factorise it versus 17 of 22 strong visualizers or 66% v 77%. Question 1, which assessed this same competency, was correctly answered by 41 of 55 weak visualizers and 50 of 60 strong visualizers or 75% v 83%. Most of the attrition occurred before the solution phase in which the core competency was needed.

	All 4 ingredients	Correct equation	Solve equation
Weak visualizers	36%, n = 20	16%, n = 9	11%, n = 6
Strong visualizers	53%, n = 32	37%, n = 22	28%, n = 17

Table 6-5. Distribution of weak and strong visualizers among problem solving codes for the Lawn Problem.

Success on the Lawn problem revealed a medium to large effect size in spatial ability ($d = .54$, $p < .05$) and a very similar result for the corresponding competency question ($d = .56$, $p < .05$). When those who responded incorrectly to the competency question were excluded the effect size was almost unchanged ($d = .52$, $p < .05$). With regard to the problem solutions, large effect sizes were observed for codes related to translating the relational expressions in the problem statement (codes 2 & 3 above) and selecting area as the appropriate schema for the problem (code 6). Strong visualizers were more likely than weak visualizers to be successful in gathering the ingredients needed to write the quadratic equation and to then write this equation.

6.2 Coding of the remaining problems

The same style of approach outlined for the Lawn problem was then followed for the other five problems – apply Mayer's framework to each problem to develop a set of actions, begin to analyse the data by scoring solutions against this list, edit the list itself as new actions were found and ensure all solutions were checked for all actions.

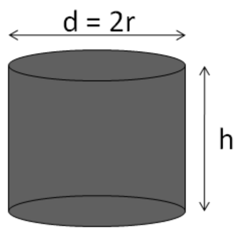
The Jug Problem

The Jug Problem is provided below:

Stainless steel cylindrical jugs are made to hold a volume of 2 litres (2000 cm3). If the 1 litre mark is at 8.84 cm what is the radius of the jug to the nearest centimetre?

The core competency question related to this problem is Question 2, provide the equation for the volume of a cylinder:

What is the volume of this cylinder?



Identification of the different knowledge components in the problem based on the Mayer schema led to the following a priori codes:

- Linguistic knowledge is required to discern the total volume of 2000 cm^3 can be ignored and that half the volume, 1000 cm^3, corresponds to or is matched with a cylinder height of 8.84 cm (Match 1),
- Semantic knowledge is required to understand what is meant by a jug and to recall the equation for the volume of a cylinder (the problem states the jug has a cylindrical shape),
- Schematic knowledge is required to decide this is a volume problem in which the container is a cylinder so the volume provided is represented as a cylinder volume,
- Strategic knowledge is manifest in the way the solution steps are organised: the volume provided is first equated with the cylinder volume term and this equation is rearranged to solve for one unknown, the radius, and
- Procedural knowledge is required to rearrange terms and employ arithmetic.

During the data analysis the following codes were added:

- A second match between height and volume of 17.68 cm (2 x 8.84) and 2 l or 2000 cm^3 (Match 2).
- Since it was sensible to solve the problem using either combination of height and volume the presence of either code was contained in a code called Match1 or Match2.
- Creating a sketch during problem solving.

- Including a jug height on a sketch.
- Working correctly with units for height and volume.

The full list of codes used to analyse the data is provided in Table 6-6.

No.	Component	Absent if …. (0)	Present if…. (1)	Type of knowledge	Required to solve problem
1	Match1	Not present	h=8.84 & V=1 l	Schematic and linguistic	Yes (or code 2)
2	Match2	Not present	h=17.68 & V=2	Schematic and linguistic	Yes (or code 1)
3	Matches h with V	Not present	Matches a h with a V	Schematic	Yes
4	Correct cylinder volume	Not present	Uses correct equation for cylinder volume in problem	Semantic	Yes
5	Equates volumes	Not present	Equates volume provided with volume of cylinder	Schematic	Yes
6	Sketch	Not present	Includes a sketch	Strategic	No
7	Height on Sketch	Not present	Indicates a height mark on sketch	Strategic	No
8	Units	Not present	Uses correct units, e.g. 1l = 1000 cm^3	Semantic	Yes
9	Match1 or Match2	Not present	Either match from code 1 or code 2	Schematic and linguistic	Yes

Table 6-6. Coding scheme for the Jug Problem

A summary for all codes is shown in Table 6-7 below along with a comparison of spatial test scores for those demonstrating and not demonstrating each code in their solutions.

			Weak n		Strong n		Total n		PSVT:R			
No.	**Code**	**1=**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**t (p)**	**d**
1	Match1	h=8.84&V=1	24	31	33	27	57	58	21.40 (5.00)	18.03 (5.72)	-3.360 (.001)**	0.63 (Large)
2	Match2	h=17.68&V=2	15	40	20	40	35	80	19.57 (5.30)	19.76 (5.78)	.353 (.867)	0.04 (Small)
3	Matches h with V	h matched with a V	49	6	58	2	107	8	20.02 (5.43)	15.50 (6.99)	-2.234 (.027)*	0.66 (Large)
4	Correct cylinder volume	Correct vol eqn	35	20	53	7	88	27	20.70 (5.24)	16.44 (5.65)	-3.269 (.000)**	0.79 (Large)
5	Equates volumes	Vol provided equated with cyl vol	45	10	55	5	100	15	20.14 (5.44)	16.80 (6.11)	-2.184 (.031)*	0.58 (Large)
6	Sketch	Sketch included	35	20	39	21	74	41	19.80 (5.67)	19.54 (5.58)	-.238 (.813)	0.05 (Small)
7	Height on Sketch	Height indicated	28	27	35	25	63	52	19.94 (5.43)	19.42 (5.87)	-.487 (.628)	0.10 (Small)
8	Units	Correct units	49	6	55	5	104	11	19.99 (5.67)	17.00 (4.36)	-1.694 (.093)	0.60 (Large)
9	Match1 or Match2	Either match	39	16	53	7	92	23	20.71 (5.17)	15.70 (5.64)	-4.084 (.000)**	0.93 (Large)
10	Codes 4 AND 9	Vol. eqn. & Match 1 or 2	28	27	49	11	77	38	21.23 (4.88)	16.61 (5.78)	-4.496** (.000)	0.87 (Large)

* Significant at the p < .05 level
** Significant at the p < .01 level

Table 6-7. Statistical data for the codes in the Jug Problem

While the sample was split into those who provided a sketch and those who didn't, a very small effect size was observed in this code, i.e. the decision to sketch did not vary with spatial ability, weak visualizers were as likely or unlikely to sketch as strong visualizers. Including a height on the sketch increased its relevancy to solving the problem but this also did not correlate with spatial ability. Match2 also revealed a very small effect size. This was arguably the less obvious of the two matches between height and volume that could be made.

Equating the volume provided with a cylinder volume, code 5, was essential in solving the problem and, with a medium to strong effect size, this was somewhat more likely to be found in the solutions of strong than weak visualizers. A small number of participants, 11, made errors in units, with a slightly higher proportion of these being weak visualizers. While almost all participants matched a height with a volume the sample was evenly split by the Match1 code, h = 8.84 cm and V = 1000 cm^3, with this match more likely to be made as spatial ability increased. As mentioned, the problem could equally well be solved using Match2 but of the 23 who failed to make either appropriate match between height and volume, 16 were weak visualizers and a large effect size, $d = .93$, $p < .001$, was observed on this code. Some participants either didn't use the cylinder volume expression or used an incorrect version in their solutions and these were more likely to be weak visualizers (20 out of 27).

Jug problem summary

The Jug problem revealed a large and highly significant difference in spatial ability with 67 participants being correct and 46 incorrect ($d = .92$, $p < .001$). Recall of the correct cylinder volume expression also revealed a difference in spatial ability with an effect size, $d = .57$ ($p < .05$). Excluding those who did not know the cylinder volume expression, a slightly higher effect size was observed between correct and incorrect groups on the problem ($d = 1.09$, $p < .000$). This was a relatively simple problem that required a correct match between jug height and volume in order to calculate radius. Strong visualizers were more likely than weak visualizers to make an appropriate match but the use of sketching to facilitate the problem solution was favoured to the same extent by both groups.

The Cans Problem

The Cans Problem is provided below:

> *Drink cans are made by stamping out circular discs from a sheet of metal. The rectangular sheet from which the discs are stamped out measures 1 m by 2 m. If the cans have a radius of 10 cm, how many discs can be made from this sheet of metal?*

There was no core competency question related to this problem. Although the discs are circles, neither the area nor the circumference are needed in the correct solution.

A priori codes based on the Mayer schema.

- Linguistic knowledge is required to discern that a thin sheet of metal has a rectangular shape whose dimensions are 1 x 2 m, the discs are circles and the discs are being cut out of the sheet.
- Semantic knowledge is required to understand that drink cans are cylinders, a radius is a property of a circle and is half the diameter and stamping a round shape from a rectangular sheet will lead to waste. Semantic knowledge is also required to understand that discs are circles but this was stated in the problem.
- Schematic knowledge is required to determine the pattern of stamping is a grid formed by dividing the sheet into rows and columns, the width of each equal to the diameter of the circle
- Strategic knowledge is manifest in the way each schema is implemented.
- Procedural knowledge is required to complete the arithmetic required in the solution.

A posteriori codes developed during the data analysis

- Variations on grid schema: while the correct grid is based on diameter to give a 5 x 10 grid, several errors were made in calculating the number of rows and columns. The most common error was to use the radius to give a 10 x 20 grid but there were several other inappropriate combinations which were grouped together as 'grid = other'. This also led to the creation of a code called grid scored as 1 for any type of row x column grid.
- Additional schema no. 1: use a staggered pattern when stamping the circles to reduce waste and produce more cans per sheet than with the grid pattern
- Additional schema no. 2: assume there is no waste, that all of the sheet area can be converted to can area, number of cans is then sheet area divided by can area (area of circle)

- Additional schema no. 3: assume there is no waste but divide the sheet area by a linear property of the disc – circumference, radius or diameter. This contains a dimensional error since the result of dividing an area ($length^2$) by a length is a length and not a number of discs.

A complete list of codes is shown in Table 6-8 along with a comparison of spatial test scores for the correct versus incorrect groups.

			Weak n		**Strong n**		**Total n**		**PSVT:R**			
No.	**Code**	**=1 if**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**t (p)**	**d**
	Sheet shape	Rectangle	54	1	60	0	114	1				
	Sheet dimensions	1 x 2 m	54	1	59	1	113	2				
1	Diameter	Calculated	18	37	35	25	53	62	21.06 (4.81)	18.55 (6.02)	-2.440* (.016)	0.47 (Medium)
2	Grid 1	5x10	12	43	33	27	45	70	22.16 (4.73)	18.13 (5.60)	-3.993** (.000)	0.78 (Large)
3	Grid 2	10x20	5	50	6	54	11	104	20.09 (6.44)	19.66 (5.55)	-.239 (.811)	0.08 (Small)
4	Grid 3	Staggered rows	0	55	2	58	2	113	25.50 (3.53)	19.60 (5.60)	N/A	N/A
5	Grid 4	Other dimension grid	5	50	5	55	10	105	18.80 (6.00)	19.79 (5.60)	.532 (.596)	0.18 (Small)
6	Grid	Any Grid	22	33	46	19	68	47	21.43 (5.30)	17.21 (5.14)	-4.242** (.000)	0.81 (Large)
7	Schema 2	Sheet Area/ Disc Area	20	35	9	51	29	86	17.31 (5.61)	20.51 (5.41)	2.730** (.007)	0.59 (Large)
8	Schema 3a	Sheet Area/ Diameter	5	50	1	59	6	109	16.17 (4.17)	19.90 (5.63)	1.596 (.113)	0.76 (Large)
9	Schema 3b	Sheet Area/ Circumf.	2	53	2	58	4	111	19.75 (6.70)	19.70 (5.61	-.016 (.987)	0.01 (Small)
10	Schema 3c	Sheet Area/ Radius	3	52	0	60	3	112	16.67 (1.16)	19.79 (5.66)	.949 (.344)	0.77 (Large)
11	Schema 3	Sheet Area/Linear	10	45	3	57	13	102	17.38 (4.63)	20.00 (5.68)	1.593 (.114)	0.51 (Large)

* Significant at the p < .05 level
** Significant at the p < .01 level

Table 6-8. Statistical results for each code in the Cans Problem.

The smallest effect sizes were observed with codes 3, 5 and 9. For each, the membership of one of the categories, code =1, was n = 11 or lower and, therefore, too small to form any conclusions. Rather than look at them separately, each of these codes is contained within another: 3 and 5 are contained in code 6, 'Any Grid', and 9 is contained within 11, 'Sheet Area/Disc Linear', and, hence, can be discussed under these codes.

Of medium effect size, the 'Diameter' code was more likely to be found in the solutions of strong visualizers. A participant was scored as 1 for this code if his/her solution merely contained the diameter value (20 cm or 0.2 m) somewhere in the solution regardless of presenting it as a formal calculation from radius or not. This value was essential in obtaining the correct grid dimensions and, hence, the correct answer to the problem.

Codes 8 and 10, which revealed large effect sizes and low membership in the 1 category, were not only a misrepresentation of the problem but also lacked sense. When grouped with code 9 to form 'Sheet Area/Linear', the membership was only slightly higher ($n = 13$) and the code revealed a moderate to large effect size in favour of weak visualizers indicating that weak visualizers were somewhat more likely to take this solution approach.

While dividing the sheet area by disc area was sensible at a dimensional level it too was a misrepresentation of the problem as there was no reference to recycling of material in the problem statement. A moderate to large effect size was found for this code (no. 7) with weak visualizers more likely to adopt this than strong visualizers.

Representing the sheet as a 5 x 10 grid, code 2, was essential to getting the correct answer to the problem. The sample was split almost in half by this code with a large effect size in favour of strong visualizers. When grouped with all other grid representations, the sample was again split near the middle and again in favour of strong visualizers with an effect size of .81, $p < .001$. Lastly, the arguably more advanced schema of staggered rows was in evidence in only two responses which happened to come from participants with higher than average spatial scores but this number is too low to rule out this occurring by chance. Neither managed to solve the problem in this way.

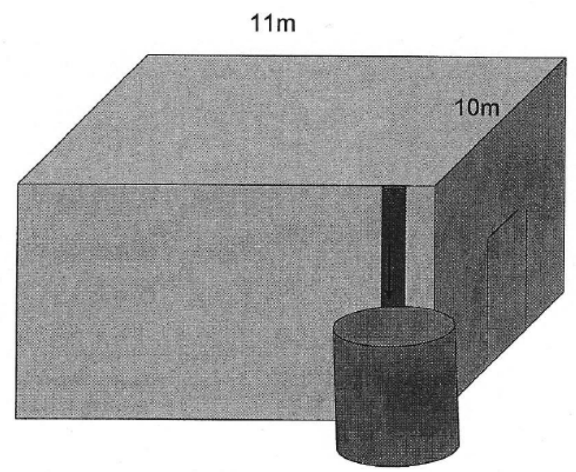
Cans problem summary

For the Cans problem, 44 participants provided the correct answer and 71 did not. The difference in spatial ability of these two groups produced a large and very significant effect size ($d = .81$, $p < .001$). There was no core competency question associated with this problem. Three representations of the problem were in evidence in the data: (i) a grid layout that resulted in waste material, (ii) entire sheet converted to cans based on area ratio and (iii) an impossible division of sheet area by a linear property of the disc that appears to be accompanied by zero waste material. A large and highly significant effect size ($d = .81$, $p < .001$) in favour of strong visualizers was observed for the first representation while moderate to large effect sizes in favour of weak visualizers were measured for the second two representations ($d = .59$, $p < .01$ and $d = .51$, N.S., respectively).

The Rain Problem

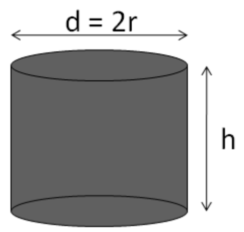
The Rain Problem is provided below:

"*The diagram shows the dimensions of a flat roofed commercial shed. During one week 5 mm of rain fell on the roof of the shed. The rain was collected by gutters that flowed into a cylindrical water barrel with a diameter of 1 m. By how much did the depth of the water in the barrel increase as a result of this rain?*"

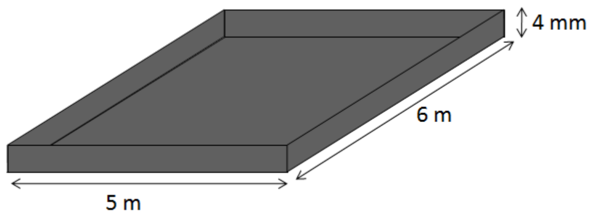


There are two core competency questions, 2 and 5, associated with this problem and these are, respectively:

Q2. What is the volume of this cylinder?



Q5. What is the volume of this tank?



Identification of the different knowledge components in the problem based on the Mayer schema.

- Linguistic knowledge is required to discern the roof of the commercial shed surrounded by gutters is reduced to a rectangle for the purpose of this problem and that rain is water that flows from the roof to the barrel,
- Semantic knowledge is required to understand that gutters collect the rain from the roof and channel it to the barrel. It is also required to convert units of length from mm to m or vice versa. Recall of the cylinder volume expression and the prism volume expression are also required.
- Schematic knowledge is required to discern the rain fall as creating a third dimension of height to turn the 2-D roof rectangle into a 3-D roof prism; the volume of this prism is then completely transferred (i.e. volume is conserved) to the cylindrical barrel. Hence, the schema for this problem contains 3 parts:
    - The roof is a rectangular prism the height of which is the amount of rain fall
    - This volume is transferred to the barrel and the conservation of volume is maintained

- The barrel is a cylinder (either sketch indicates this or cylinder volume expression is provided). This shape is stated in the problem.

- Strategic knowledge is manifest in the decision to calculate the rain volume first, equate this to the barrel volume and then solve for the unknown barrel height.
- Procedural knowledge is required to complete the arithmetic required in the solution.

Codes derived from the data include steps, actions or decisions made by the participants and were found to include:

- An additional schema (inappropriate) that 5 mm of rain is a property of the barrel, manifest in the following ways:
  - the answer to the problem (i.e. increase in barrel level) is 5 mm
  - the barrel volume is set to 5 mm and a height calculation is then attempted
  - the barrel volume is calculated using a height of 5mm.

Each code was then used to group the sample to compare PSVT:R scores of those demonstrating the code in the problem solution to those who didn't. These results are included in Table 6-9

			Weak n		Strong n		Total n		PSVT:R			
No.	Code	1 =	1	0	1	0	1	0	1	0	t (p)	d
	Roof shape	Discerns roof to be 2-D rectangle	35	21	49	11	84	32	20.51 (5.41)	17.28 (5.34)	-2.830** (.006)	0.61 (Large)
	Roof units	5 mm = 0.005 m	32	24	39	21	71	45	20.70 (5.30)	17.91 (5.84)	-2.657** (.009)	0.51 (Large)
	Rainfall amount	= 5mm	51	5	58	2	109	7	19.92 (5.58)	15.00 (5.23)	-2.268 (.025)	0.91 (Large)
	Barrel diameter	= 1m	48	8	54	6	102	14	19.75 (5.60)	18.71 (6.22)	-.637 (.525)	0.18 (Small)
	Cylinder expression	Correct expr for cylinder volume	25	21	46	8	71	29	21.13 (5.33)	16.14 (5.03)	-4.315** (.000)	0.97 (Large)
	Rainfall= prism height	5mm=height of prism	27	29	44	16	71	45	21.20 (5.32)	17.13 (5.32)	-4.007** (.000)	0.77 (Large)
	Volume equation	V=LxWxH	26	30	46	14	72	44	21.26 (5.27)	16.93 (5.291)	-4.291** (.000)	0.83 (Large)
	Volume conserved	Vol. transferred & conserved	23	33	44	16	67	49	21.57 (5.28)	16.96 (5.10)	-4.713** (.000)	0.89 (Large)
	Barrel shape	Cylinder (vol. expr or sketch)	50	6	57	3	107	9	19.89 (5.53)	16.44 (6.54)	-1.769 (.080)	0.57 (Large)
	Solution order	Direct order	16	40	43	17	59	57	22.29 (5.17)	16.86 (4.78)	-5.865** (.000)	1.10 (Large)
	Correct roof volume	Volume = 0.55 m^3	19	37	36	24	55	61	21.71 (5.02)	17.74 (5.58)	-4.013** (.000)	0.75 (Large)
	Rainfall= barrel height	Barrel ΔL=5mm	19	37	8	52	27	89	17.37 (4.39)	20.30 (5.84)	2.407 (.018)	0.57 (Large)

* Significant at the p < .05 level
** Significant at the p < .01 level

Table 6-9. Statistical data for the codes in the Rain Problem

All of the codes revealed large effect sizes ranging from .51 to 1.10, except for the barrel diameter code. The smallest of these effect sizes related to unit conversion with strong visualizers slightly more likely to correctly convert 5 mm to 0.005 m (d = .51, p < .01). Code 1, which required the participant to show in any way that the roof has a rectangular shape, revealed a moderate to large effect size in spatial ability (d = .61, p < .01) with weak visualizers more likely than strong visualizers (38 v 18 %) to omit this key component in the solution.

With regard to representation, two versions were discerned in the data set – rainfall amount (5 mm) creating a volume that was transferred to the barrel and rainfall amount being a property of the barrel. A large effect size (d = .89, p < .001) was measured for the former representation with strong visualizers more likely to adopt it while the latter was more likely to be adopted by weak visualizers with a moderate to large and less significant effect size measured (d = .57, p < .05).

Rain problem summary

The rain problem involved equating of two different volume expressions – cylinder and prism – and revealed a large and significant effect size in spatial ability between those who answered it correctly and those who didn't (d = .83, p < .01). Prior knowledge of the two volume expressions – cylinder and prism – was assessed in Q2 and Q5 and when participants were excluded if incorrect on both the effect size remained the same between correct and incorrect on the problem. All participants demonstrated knowledge of the prism volume expression in either the core competency question or the problem solution. Correct recall of the cylinder volume expression proved more challenging with 25 participants failing to do so in either the problem solution or the core competency question with a close to significant difference in spatial ability between them and those who knew the expression (d = .44, p = .053). A very different effect size was observed in the code related to the use of this expression in the problem solution. Of the 91 participants who knew the cylinder volume expression, only 71 included it in the solution and when their spatial test scores are compared against those who didn't use the correct expression (missing or incorrect), a much larger and more significant effect size was observed with d = ..97, p < .001.

The Pencils & Jars Problem

The Pencils and Jars Problem is provided below:

I have some pencils and some jars. If I put 4 pencils into each jar I will have one jar left over. If I put 3 pencils into each jar I will have one pencil left over.

How many pencils and how many jars are there?

The core competency question related to this problem is Question 6, solve for x and y in to simultaneous linear equations:

Determine the value of x and y by solving these two equations

$$x + y = 6$$

$$-3x + y = 2$$

Identification of the different knowledge components in the problem based on the Mayer schema are:

- Linguistic knowledge is required to discern that (i) a jar is empty with 4 pencils per jar and (ii) with 3 pencils per jar there is an extra pencil – these are relational statements – and can be translated to algebraic equations.
- Semantic knowledge is required to understand what is meant by a pencil and a jar.
- Schematic knowledge is required to decide that the two key statements in the problem can and should be converted to two equations which can be solved simultaneously to give the answer.
- Strategic knowledge is manifest in writing each equation first, then equating them to eliminate an unknown, solving for the unknown and using this value in either original equation to calculate the other unknown.
- Procedural knowledge is required to complete the arithmetic required in the solution and, in addition if modelling is used, to employ basic algebra.

When the data were analysed additional schematic knowledge codes emerged – guess and check with pictures and/or a number pattern.

- Additional schematic knowledge was evident in the use of a guess and check approach with symbols. This was evident in solutions which contained sketches of two series of U shaped jars containing either lines or a number to represent pencils. The first series would have 4 lines or the number 4 per jar while the second series would have 3 per jar. Different numbers of jars would be incrementally tested by adding jars to the existing series or creating new pairs of series for each new guess. Each guess was evaluated by referring to the 'rules' provided in the problem statement. The related strategy was to first guess and, if unsuccessful, decide whether to increment up or down for each of the subsequent guesses.

- A variation on this guess and check schema was to use a number pattern instead of sketches. The rules for the number pattern or sequence was to find a number that was divisible by 4 with no remainder and divisible by 3 with a remainder of 1.
- Understanding of the two relational statements in the problem was evident in the guess and check approach even when they were not expressed as algebraic equation. Participants were scored as 1 for each statement if correctly translated it into an equation or if they showed evidence through a sketch of jars and pencils with the correct remainders or if they showed evidence in a number pattern consistent with the statements.

A summary for most of these codes is shown in Table 6-10 along with a comparison of spatial test scores for the correct v incorrect groups.

		Weak n		**Strong n**		**Total n**		**PSVT:R**			
Code	**1=**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**t (p)**	**d**
Statement 1	Correct	39	17	29	31	68	48	19.01 (5.81)	20.48 (5.39)	1.378 (.171)	0.27 (Medium)
Statement 2	Correct	35	21	25	35	60	56	19.15 (5.75)	20.13 (5.57)	.926 (.356)	0.18 (Small)
Guess & Check	Yes	47	9	30	30	77	39	18.09 (5.46)	22.64 (4.83)	4.405** (.000)	0.89 (Large)
Sketch	Sketch used	36	20	26	34	62	54	18.42 (5.13)	21.00 (5.97)	2.505 (.014)*	0.47 (Medium)
Check with sketch	Guess & Check and sketch used	36	20	24	36	60	56	18.25 (5.06)	21.09 (5.94)	2.777 (.006)	0.52 (Large)
Number Sequence	Guess with num sequence	8	48	6	54	14	102	19.36 (5.34)	19.66 (5.73)	.185 (.854)	0.06 (Small)
Model Attempted	Yes	21	35	33	27	54	62	21.09 (5.99)	18.34 (5.06)	-2.683 (.008)**	0.50 (Large)

* Significant at the $p < .05$ level
** Significant at the $p < .01$ level

Table 6-10. Statistical data for the codes in the Pencils & Jars Problem

The use of two simultaneous equations to solve this problem reflects a decision by the participant to use a mathematically based approach and was more likely to be taken by strong than weak visualizers (55 % v 38 %). Those who chose to model were scored on their success in doing so as having 0, 1 or 2 correct equations as shown in Table 6-11. A correlation between the number of correct equations and the PSVT:R scores for these 54 participants was measured to be r = .276, p < .05. This indicates some variation shared between the two measures but membership of the 1 and 2 correct equations categories is very small. More data would need to be collected to examine any potential relationship in detail.

No. of correct equations	N	Mean	Std. Deviation
0	45	20.78	5.604
1	5	18.60	8.649
2	4	27.75	1.708
Total	54	21.09	5.991

Table 6-11. Success in translating statements to algebraic equations.

The alternative approach to solving the problem, guess and check, arguably lacks any coherent schema – it is schema free. This code revealed the largest and most significant effect size for this problem and was in favour of weak visualizers, i.e. they were more likely to adopt this approach in comparison with strong visualizers (84 % v 50 %). Two different ways of implementing guess and check were evident in the sample – check by sketching and number sequence. The former was more popular – of the 77 participants who used Guess & Check, 14 used a number sequence while 62 included a sketch (and 2 did neither).

Pencils & Jars problem summary

For the Pencils and Jars problem, 43 participants provided the correct answer and the average PSVT:R score of this group was 19.19 (S.D. = 5.24). 72 did not provide a correct response, including 4 who did not attempt it, and the average spatial test score for this group was 20.01 (5.84). The difference in these two means is very small and insignificant (d = .15, N.S.). With regard to the core competency question, 96 participants were able to answer this correctly and 19 did not. The average spatial test scores of these two groups were 19.76 (5.70) and 19.22 (5.99) and the significance value of the t-test was p = .718. Including the core competency question reflected an assumption that this problem would be solved algebraically.

Those who took a guess and check approach did not require this competency. Hence, it does not make sense to follow the procedure used in the other problem analyses of excluding those who were incorrect on the core competency question and recalculating the effect size. This is the only problem that did not reveal a difference in spatial ability levels between those who solved it correctly versus those who did not solve it correctly. However, two different approaches to solving the problem were observed – a guess and check approach favoured by weak visualizers and an algebraic approach favoured more by strong visualizers.

The Blood problem

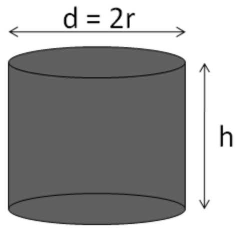
The Blood Problem is provided below:

When blood samples are centrifuged the blood separates into two distinct layers – one made up mainly of plasma and the other made up of red blood cells.

A sample of blood was put in a flat bottomed test tube with a diameter of 3 cm. When the blood sample was added to the tube it filled the tube to a depth of 7.5 cm. After centrifuging, the red blood layer was 1.5 cm deep. What volume of plasma was in the sample?

The core competency question related to this problem is Question 2, provide the equation for the volume of a cylinder:

What is the volume of this cylinder?



Identification of the different knowledge components in the problem based on the Mayer schema are:

- Linguistic knowledge is required to discern the relational statement between total, blood and plasma level/height and extract this information from the description of a mixture of two components being separated.
- Semantic knowledge is required to understand that a test tube is a small glass cylinder and to recall the equation for the volume of a cylinder (this is hinted at in the problem by providing a diameter) and to recall radius is half the diameter
- Schematic knowledge is required to decide this is a volume problem in which the container is a cylinder that contains two volumes – plasma and red blood cells - and the unknown volume is represented as a cylinder volume (Schema 1)
- Strategic knowledge is manifest in the way the solution steps are organised. Perhaps the most direct approach is to subtract heights and calculate the volume.
- Procedural knowledge is required to rearrange terms and employ arithmetic.

Codes derived from the data include steps, actions or decisions made by the participants and were found to include:

- Additional strategic action of creating a sketch during problem solution
- Complementing the sketch by indicating a height
- Additional strategic approach is to calculate two volumes – red blood cells and total – and subtract to give plasma volume
- Additional schema of seeing volume as height, i.e. the answer to the problem is given as a height (Schema 2)
- Additional schema of seeing volume as area, i.e. the answer to the problem is given as a vertical cross sectional area of the test tube (Schema 3)
- Additional schema of seeing container as being a prism, i.e. the answer to the problem is given as the volume of a prism rather than a cylinder (Schema 4)

All codes for this problem are shown in Table 6-12.

No.	Component	Absent if …. (0)	Present if…. (1)	Type of knowledge	Required to solve problem
1	Height subtracted	No heights subtracted	Plasma height obtained from subtracting 2 heights	Linguistic and/or schematic	No if volume subtracted
2	Volume expression	Not provided or incorrect	Correct cylinder volume expression	Semantic	Yes
3	Radius	Radius not calculated	Radius is calculated	Semantic	No
4	Schema 1 Volume		Plasma volume is the volume of a cylinder	Schematic	Yes
5	Plasma volume	A plasma volume not calculated	A correct or incorrect plasma volume calculated	Strategic	Yes
6	Sketch	Not present	Includes a sketch	Strategic	No
7	Height on Sketch	Not present	Indicates a height mark on sketch	Strategic	No
8	Red cells volume	A red cells volume not calculated	A correct or incorrect volume calculated	Strategic	No
9	Total volume	A total volume not calculated	A correct or incorrect volume calculated	Strategic	No
10	Volume subtracted	No volumes subtracted	Plasma volume obtained from subtracting 2 volumes	Strategic	No if height subtracted
11	Schema 2 Height	Not present	Plasma volume is the height of plasms	Schematic	No
12	Schema 3 Area	Not present	Plasma volume is the area of plasma	Schematic	No
13	Schema 4 Prism	Not present	Plasma volume is the volume of a prism	Schematic	No

Table 6-12. Coding scheme for the Blood Problem

A summary for most of these codes is shown in Table 6-13along with a comparison of spatial test scores for the correct versus incorrect groups.

			Weak n		Strong n		Total n		PSVT:R			
No.	Code	1=	1	0	1	0	1	0	1	0	t (p)	d
1	Height subtracted	Plasma vol obtained by height subtract	33	22	34	26	67	48	19.90 (5.19)	19.44 (6.21)	-.430 (.668)	0.09 (Small)
2	Volume expression	Correct cylinder volume expr.	26	29	47	13	73	42	21.34 (4.98)	16.86 (5.60)	-4.454** (.000)	0.85 (Large)
3	Radius	Radius computed	34	21	53	7	87	28	20.91 (5.15)	15.96 (5.42)	-4.362** (.000)	0.94 (Large)
4	Schema 1	Representation= cylinder vol	38	17	54	6	92	23	20.75 (5.19)	15.52 (5.38)	-4.289** (.000)	0.99 (Large)
5	Plasma volume	A plasma vol provided	41	14	50	10	91	24	20.48 (5.27)	16.75 (6.02)	-2.998** (.003)	0.66 (Large)
6	Sketch	Sketch included	38	17	38	22	76	39	19.49 (5.16)	20.13 (6.46)	.132 (.564)	0.11 (Small)
7	Height on sketch	Height included	34	21	33	27	67	48	19.46 (5.24)	20.04 (6.14)	.544 (.588)	0.11 (Small)
8	Red cells volume	A red cells vol provided	16	39	20	40	36	79	20.00 (5.76)	19.57 (5.58)	-.380 (.705)	0.08 (Small)
9	Total volume	A total vol provided	16	39	21	39	37	78	20.49 (5.74)	19.33 (5.55)	-1.029 (.305)	0.21 (Medium)
10	Volume subtracted	Plasma vol obtained by vol subtract	15	40	19	41	34	81	20.38 (5.53)	19.42 (5.66)	-.838 (.404)	0.18 (Small)
11	Schema 2	Representation= height diff	6	49	3	57	9	106	16.33 (3.97)	19.99 (5.65)	1.898 (.060)	0.75 (Large)
12	Schema 3	Representation= area calc	2	53	1	59	3	112	N/A	N/A	N/A	N/A
13	Schema 4	Representation= prism vol	1	54	0	60	1	114	N/A	N/A	N/A	N/A
Excluding those who did not attempt the problem, 4 weak and 3 strong												
	Volume = Cylinder	Representation=cylinder vol	31	11	61	5	92	16	20.75 (5.19)	15.50 (4.02)	-3.844** (.000)	1.14 (Large)
	Answer	Correct answer	16	26	46	20	62	46	21.76 (4.71)	17.57 (5.28)	-4.345** (.000)	

* Significant at the $p < .05$ level

** Significant at the $p < .01$ level

Table 6-13. Statistical data for the codes in the Blood Problem

As for the Jug problem, sketching as part of the solution did not reveal a difference in spatial ability. Small and insignificant effect sizes were also found for the strategy codes: code 1 indicated a solution that first subtracted heights and then calculated a volume which code 10 indicated that volumes were first calculated and then subtracted. Both approaches led to the correct answer and neither was more favoured by one group of students than the other.

Not all participants calculated a radius and this code revealed a large and significant effect size. The largest and most significant differences emerged on the schema codes. Those who represented the problem as a cylinder volume calculation for which height and radius are known, i.e. the correct schema, had a significantly higher spatial test score than those who didn't ($d = .99$, $p < .001$). Of the 23 who failed to show this schema, 7 did not attempt the problem and when removed the effect size increased ($d = 1.14$, $p < .001$). The remaining 16 participants who failed to show this schema were divided in the alternative schema they selected with membership of each schema code being too small to draw any meaningful conclusions through statistical comparisons.

This problem is similar to the Jug Problem. Both are cylinder volume problems that require simple application of the cylinder volume equation to find one unknown from height, diameter and volume. In this case, the unknown is the volume with the height and diameter contained in the problem statement. Participants who got this correct typically solved the problem by either calculating and then subtracting total and red cells volumes or first subtracting total and red cells heights before calculating one volume, that of the plasma.

Blood problem summary

For the Blood problem, 62 participants provided the correct answer and the average PSVT:R score of this group was 21.76 (S.D. = 4.71). 53 did not provide a correct response, including 7 who did not attempt it, and the average spatial test score for this group was 17.30 (5.67). As measured by an independent samples t-test, the difference in these two means was found to be significant with $t = -4.604$, $p < .001$ and Cohen's $d = .86$.

With regard to the core competency question, the average spatial test score of those who correctly wrote the cylinder volume equation was significantly higher than the average spatial ability of those who did not ($n = 86$ v 29, PSVT:R = 20.48 v 17.41, $t = -2.605$, $p \leq .01$).

Conclusions

Taking the data collected from each problem in turn, participants' solutions were analysed to learn why strong visualizers were more successful than weak visualizers in both the representation and solution phases. While the Mayer schema of mathematical problem solving produced a list of codes that resonated well with the participants' problem solving actions, some additional and qualitatively different codes emerged during the analysis and were added to the list. In some cases, codes were evident in the majority of solutions and did not discriminate by spatial ability. An example is the use of sketching in the Jug and Blood Problems. Other codes revealed moderate effect sizes such as calculating a diameter in the Jug Problem and having all four ingredients to write the equation for the Lawn Problem. Among the largest effect sizes were those related to problem representation codes such as the grid representation on the Cans Problem, Area = W x L on the Lawn Problem and making an appropriate match between h and V on the Jug Problem. These codes covered different aspects of problem representation from linguistically comprehending the contents of the problem statement to drawing on or recalling semantic knowledge and selecting a schema on which to base the solution step.

While the research thus far had revealed some important findings, several questions still remained. To what extent were participants consistently successful or unsuccessful at the linguistic and schematic levels? Why was the Pencils and Jars problem exceptional? What were the characteristics of this problem that made it different from the others? What characteristics must a problem have in order for it to differentiate between weak and strong visualizers? Were the linguistic and schematic knowledge terms, adopted from Mayer (1992), the best way to describe the differences in thinking evident in the problem solutions? A third phase of analysis of the data was now required which looked across the problem solutions to determine the variation in the ways the problems were interpreted and represented.

Chapter 7 Variation in representation across problems and across participants

Introduction

Problem representation is a complex, challenging process that draws on several types of knowledge and places high demands on cognitive attention, especially when problems are novel or non-routine. Despite being asked to think aloud, participants in this research project would often go quiet as they strained to connect with the problem and search for its secrets. Often, a participant would read a problem out loud only to then read it again in silence as if the act of reading out loud was draining much needed cognitive resources away from the process of understanding the problem. Despite their apparent simplicity, participants failed at times to find the key to unlocking a problem. Sometimes, problems were misrepresented and, in some instances, to the extent that a different problem to the one stated was solved (e.g. a rectangular rather than square lawn, a can stamping process with recycling rather than without). While it is possible the close proximity of the researcher as interviewer affected performance of participants in the OSU sample, misrepresentations were evident in both groups. Success in problem representation was found to significantly correlate with spatial ability. Several problem representation codes revealed significant and sizeable differences in PSVT:R scores but not always so. Looking across the problems, there was variation in magnitude and type of representation code - linguistic, semantic and schematic – that correlated with spatial ability.

The purpose of this chapter is twofold. First, to analyse the way each problem was represented within the sample in order to understand the essential features of these representations and identify a key representation code from each problem. Second, these essential features are then compared across the data set by searching for similarities and differences across the problems and by examining consistency at problem representation by participant. Consistency among participants is measured to (i) determine the extent to which it varies with spatial ability and (ii) assess the extent to which success in problem representation is repeated. Comparison across problems is measured to reveal the reasons why there is variation in effect size among the problems.

7.1 Problem representation and schema development

In this section, I summarise for each problem the linguistic, semantic and schematic codes that together describe the various aspects of problem representation. I then check the extent to which each code revealed a difference in spatial ability scores in order to highlight codes that did and did not split the sample, i.e. were easy or difficult, and for those that did split the sample, check whether they split the sample by spatial ability or did not. For each problem, this leads to some categories of variation in approach to problem solving which I then try to rationalise before concluding with a key code for that problem that has three properties:

1. It was essential in representing the problem and the problem could not be solved without this representation;
2. It split the sample into two large portions;
3. This split revealed a difference in spatial ability.

Since a correlation between success and spatial ability has been shown for all but one of these problems, the objective now became one of seeking to measure the relationship between representation and spatial ability. In this section, the separate analysis of each problem is continued until a key code or aspect of representation is extracted. In the following section, these codes are then combined to form a set which is examined for common themes and differences in order to learn what these reveal about the spatial – problem solving relationship.

The Lawn Problem

> *"A **square** lawn was extended in width by 2 m and in length by 3 m. The area of the new lawn is twice as big as the area of the old lawn. What are the measurements of the old lawn?"*

As shown in Table 7-1, weak visualizers were lost at higher rates than strong visualizers at each step of representing and solving to the Lawn problem. The problem statement contains four ingredients that are required to write the equation; three are linguistic, they are contained in the words of the problem – the lawn is square, width + 2 & length + 3, and A_{new} = 2 x A_{old} - and one is a schema selected by the participant – area = width x length. Of the three linguistic codes, the 'lawn is square' had the lowest success rate for both weak and strong visualizers but did not correlate significantly with spatial ability. More weak visualizers made this omission but this code did not reveal a significant difference in spatial ability. What did reveal a difference was the summation of these codes, i.e. demonstrating all three revealed a significant difference in spatial ability. In contrast, this problem did not present a difficulty at the schematic level with most weak and strong visualizers applying the area schema to the problem – all but 8 participants showed evidence of width x length in their solutions. Hence, there was little difference between the number who had the set of three linguistic items and the set who had all four ingredients (three linguistic and the area schema). Writing the correct equation proved to be a challenge with many failing to do so even though they were on the right track having identified all four ingredients. Of those eligible to write the correct equation, i.e. successfully demonstrated Action 2 in Table 7-1, , 44 % (reduction from 36% to 16 %) of weak visualizers completed this step compared with 70 % of strong visualizers, i.e. this step revealed a large difference in success rates between weak and strong visualizers For those who made it as far as having the correct equation, the success rate on making it through the next step, solving the equation (a core competency), was 69 % for weak and 76 % for strong visualizers, i.e. not that different. At this stage, however, only 11 % of all weak visualizers and 28 % of strong visualizers remained with a large and significant difference in the spatial ability scores of these two groups. All these data are summarised in Table 7-1.

		Evidenced by (%)			
Action	**Description**	**All**	**Weak**	**Strong**	**Cohen's d**
1	Read the problem	100	100	100	-
2a	Treat lawn as square	63	56	70	.28
2b	Apply Width + 2	82	75	88	.72**
2c	Apply A_{new}=2xA_{old}	82	78	85	.64*
2	Apply 2a, 2b and 2c	47	38	55	.53**
3	Apply A = W x L	93	89	97	.93*
4	Combine 2 & 3, get all 4 ingredients	45	36	53	.54**
5	Write correct equation	27	16	37	.62**
6	Solve equation	20	11	28	.63**

* Significant at the p < .05 level
** Significant at the p < .01 level

Table 7-1. Numbers of all, weak and strong visualizers who showed evidence of Lawn problem solving actions.

With regard to the problem representation and solution phases, the issues that varied with spatial ability in the Lawn problem are listed below and subsequently discussed in turn:

1. Individually, the linguistic codes did not reveal both a significant difference in spatial ability and a division of the sample into two large groups; taken as a set, however, did result in both an almost even split and a significant difference in spatial ability,
2. There was a very high level of success in selecting the correct schema (area) and
3. There was difficulty in constructing the equation from the ingredients.

Most solutions invariably included the area schema but varied in the extent to which the linguistic codes were correctly interpreted. As shown in Figure 7-1, Participant No. 4 (P4) has correctly translated the relationship between new and old areas but has failed to treat the lawn as a square and to translate the relationships between old and new width and length.

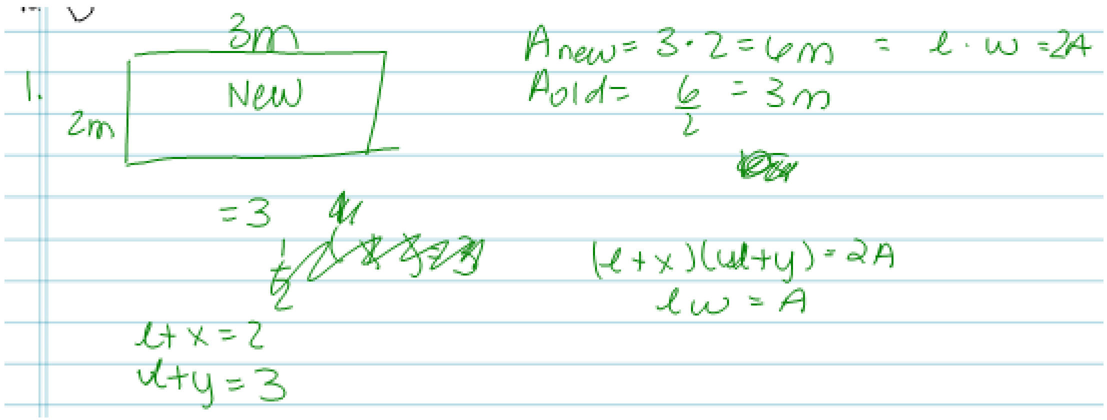


Figure 7-1. Lawn problem solution from P4 (PSVT:R = 16).

In the solution shown in Figure 7-2, P43 correctly treats the lawn as a square, translates the change in width and length and uses the area schema but makes an error in equating old and new areas. The new area is given as 4x where it should be $2x^2$. This could be a typographical error in placing the 2 as a multiplier rather than exponent of x but the second half of the solution omits the new area altogether. In fact, there is no equation in the second part, just the quadratic expression which is factorized incorrectly to give x = 2 or 3 with x = 2 selected as the correct answer. P43 does treat the lawn as square, correctly identifies the new dimensions and area schema and could be seen as correctly identifying the change in area but fails to enact this latter aspect in an appropriate way.

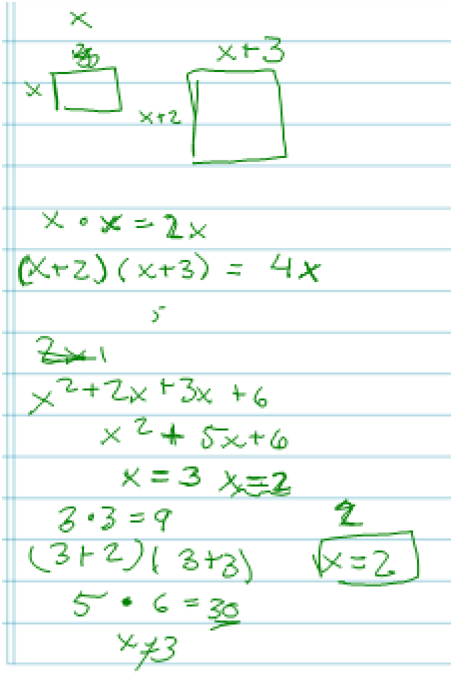


Figure 7-2. Lawn problem solution from P43 (PSVT:R = 15).

P79's first attempt to solve the problem, shown in Figure 7-3, correctly identified the area schema, the area change and the width/length changes but treated the lawn as a rectangle which led to one equation and two unknowns. P79 then began the solution again, this time treating the lawn as square. Now, P79 had all the ingredients needed to write the correct equation but failed to do so. P79 showed evidence of representing the problem correctly but failed to solve it correctly.

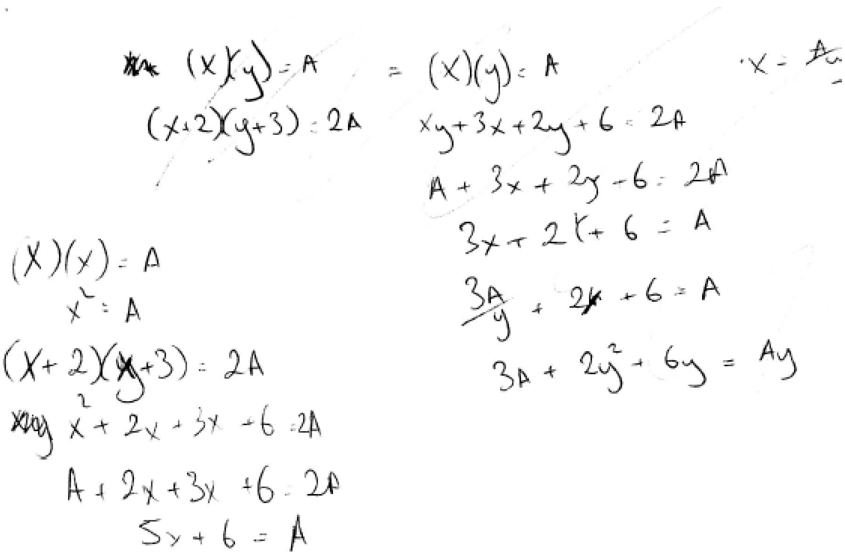


Figure 7-3. Lawn problem solution from P79 (PSVT:R = 20).

The solution by P19, in Figure 7-4, illustrates the successful identification of all four ingredients. P19 began by treating the lawn as a rectangle of width, *w*, and length, *l*, but equates them half way through the solution and replaces them each with *x* before solving the problem. P19 was successful in representation and solution.

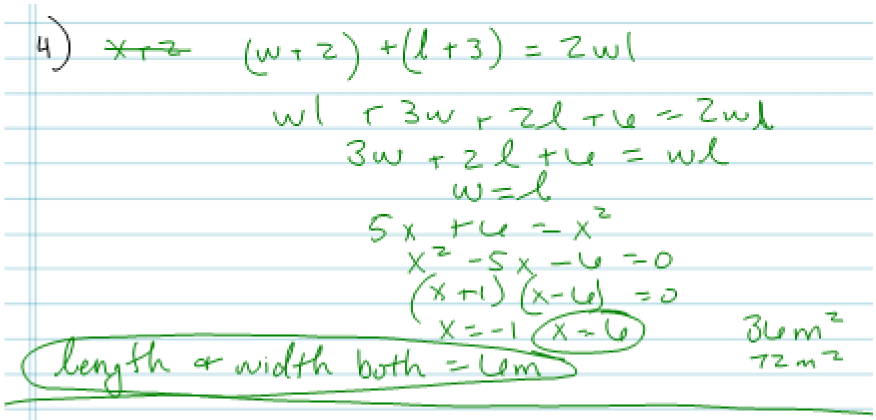
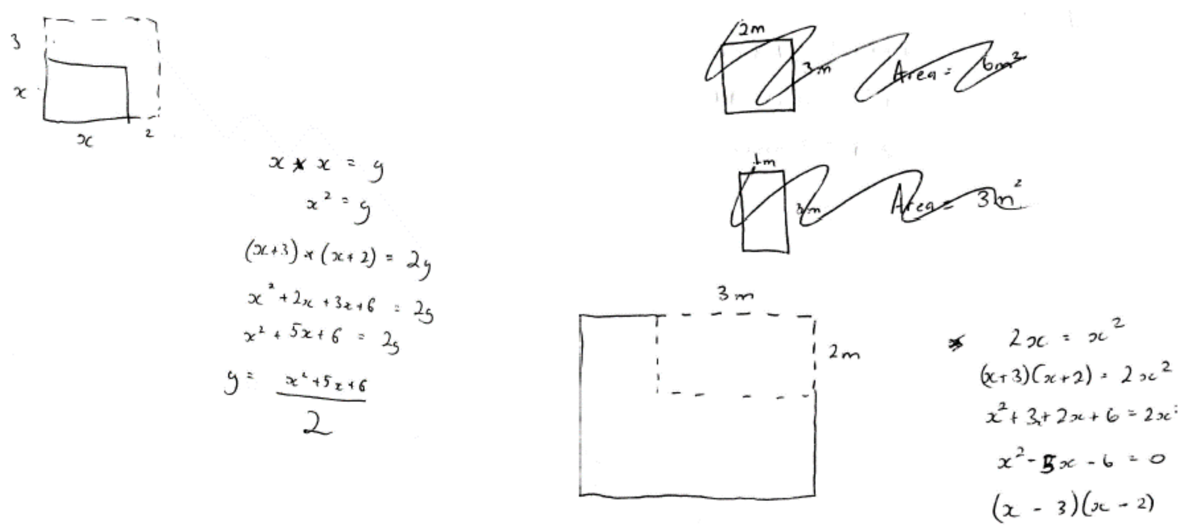


Figure 7-4. Lawn problem solution from P19 (PSVT:R = 24).

There is no reason to assume that ‘square’ is not read along with the other words in the problem statement yet, in many cases, the problem is solved as if the shape was not stated. 65 participants made this mistake and 23 corrected it (P19 being one of them), leaving 42 who did not treat the old lawn as a square. The participant must deal with four pieces of information in order to represent the problem correctly and perhaps this makes it so challenging that one or more of these items is omitted. This could reflect a working memory limitation that is connected with spatial ability. Perhaps, square is lost as it is the first to be presented and/or its loss is due to the use of the separate terms 'width' and 'length' suggesting a corresponding difference in sizes. This part of the sentence - 'In width by 2 m and in length by 3 m' - may just cue different sizes for width and length. After all, width and length are general terms that hold for any quadrilateral, including a square, and are both needed if it is not square, hence providing both may cue 'not a square' which results in 'square' being lost by a mind that is processing a lot of new information. The immediate and overt visualization of a square lawn on reading 'A square lawn …' would help prevent the loss of this piece of information or, better again, a visualization of a dynamic change in shape from square to expanded rectangle on reading the first sentence would greatly help in solving the problem. One possible explanation for the low success in identifying all three linguistic codes and in particular the square lawn may be a high demand on working memory. And, if so, to make another speculative reach, working memory may be supported by the visualization and retention of lawn shape images thereby leading to greater success rates among those with better spatial skills.

It is not surprising the area schema jumps off the page when participants read this problem given all the cues to it which include 'square', 'width', 'length', 'area' x 2, 'lawn' and 'measurements'. It should be obvious the problem is about the area of a quadrilateral and a large majority did show evidence of this schema. While 8 participants did not provide this evidence one should not conclude they were not thinking of area = width x length after they read the problem. Representation at the schematic level was not an issue in this problem.

The last issue that showed a correlation with spatial ability was the construction of the correct equation from the four ingredients with weak visualizers significantly more likely to fail at this step. The solution of P79 shown in Figure 7-3 is an example of failure to make the small extra step of writing the equation despite having everything else correct. Two other examples are shown in Figure 7-5 in which each participant comes close to writing the equation but does not provide the correct one.



(a) (b)

Figure 7-5. Lawn problem solution from (a) P60 (PSVT:R = 27) and (b) P67 (PSVT:R = 19).

The errors these participants are making relate to mathematical techniques. P60 and P79 make the same mistake of equating x^2 with another variable (y and A) but continue to use the other variable rather than the x^2. This leads to what appears to be an equation with two unknowns that can't be solved. Were they to group their variables appropriately, the quadratic equation with one unknown would have appeared and they may likely have solved the problem. P67 does make it as far as writing this quadratic equation but fails to find the factors and hence fails to solve the problem. Again, this is a mathematical error, i.e. an error in solution rather than representation. As viewed through the Mayer schema, steps that are subsequent to identifying all four ingredients are subsequent to problem representation and belong to the problem solution phase; this marks the break between problem representation and solution.

Summary of Lawn Problem representation

While strong visualizers were more likely to solve the Lawn Problem the effect size was smaller than for other problems (see Table 5-7). No single aspect of representation – linguistic, semantic or schematic – emerged as satisfying the requirements for a key code that differentiated weak and strong visualizers. The key determining aspect of problem representation in the Lawn problem is not creation of the correct equation from the four ingredients. Nor is it one of these ingredients, the area schema. Rather, it relates to the full suite of the remaining ingredients and the transfer of these three assignment and relational statements to the page in the form of symbolic or mathematical expressions, i.e. Action 2 in Table 7-1. The key code that split the sample somewhat evenly and by spatial ability is the combination of treating the lawn as square and applying width + 2 and applying $A_{new} = 2 \times A_{old.}$

The Jug Problem

Stainless steel cylindrical jugs are made to hold a volume of 2 litres (2000 cm3). If the 1 litre mark is at 8.84 cm what is the radius of the jug to the nearest centimetre?

Compared to the other problems, the Jug Problem was quite straightforward. The problem contains two short sentences which are easy to understand. The first contained the words 'cylindrical' and 'volume' which provide an obvious signpost to the schema appropriate to the problem – little thinking on the participant's part was needed to select the schema. To solve it successfully, one simply has to (i) think of a cylinder that (ii) has a height of 8.84 cm and a volume of 1000 cm^3, or a height of 17.68 and a volume of 2000 cm^3, (iii) recall the cylinder volume equation and (iv) equate the two volumes and solve for one unknown, the radius. In fact, the solution could be expressed more succinctly as: write the equation for the volume of a cylinder ($V = \pi r^2 h$), replace V with 1000, replace h with 8.84, rearrange and solve for r. It does not appear to be that problematic relative to the others with its character perhaps being closer to that of a question than a problem. The schema was contained in the problem statement and simply had to be noticed.

However, the correct answer was obtained by only 60 % of the participants with errors being made at each stage and, like the lawn problem, weak visualizers made these errors at higher rates than strong visualizers as shown in Table 7-2.

		Evidenced by (%)			
Action	**Description**	**All**	**Weak**	**Strong**	**Cohen's d**
1	Read the problem	100	100	100	-
2	Write cylinder volume expression	77	64	88	.79**
3a	Match 1 for h and V	50	44	55	.63**
3b	Match 2 for h and V	30	27	33	.04
3	Either 3a or 3b	80	71	88	.93**
4	Combine 2 and 3	67	51	82	.87**
5	Get correct answer	60	42	77	.92**

* Significant at the $p < .05$ level
** Significant at the $p < .01$ level

Table 7-2. Jug problem solving actions

This raises the following issues related to problem solving which are subsequently discussed:

1. Representation at the schematic level was correct in almost all cases,
2. Weak visualizers were less successful in recalling the cylinder volume equation,
3. The sample was evenly split in spatial ability by Match 1,
4. The arguably less direct Match 2 was equally favoured by weak and strong visualizers,
5. Of those who made either Match 1 or Match 2, weak visualizers were less likely to provide the correct expression for cylinder volume and
6. Of those who made it as far as having both the cylinder volume expression and a correct match, weak visualizers were less likely to provide the correct answer.

It was unlikely a participant would treat this problem as anything other than a cylinder volume problem as this schema was very apparent in the problem statement and this was borne out in the data. Only three participants failed to show evidence of a cylinder schema in the form of a sketch of a cylinder, expression for cylinder volume (correct or incorrect) or other reference to a cylinder. No challenge at a schematic level was evident in this problem.

The various ways jug height was matched with jug volume were as follows:

- Height not matched with a volume (inappropriate)
- h = 8.84 cm matched with V = 1000 cm^3 (or 1 litre) (appropriate)
- h = 17.68 cm matched with V = 2000 cm^3 (or 2 litres) (appropriate)
- Height mismatched with a volume, e.g. h = 8.84 matched with V = 2000 cm^3 (or 2 litres) (inappropriate)

Only 8 participants failed to match a height with a volume – one did not attempt the problem, four ignored volume values and three showed both a height and a volume but did not connect them. While this group did have a significantly lower spatial test score there are too few in this category to draw any conclusion from this. An example from this category is shown in Figure 7-6 in which P22 correctly calculates the height at the 2 litre mark as 17.68 but does not connect it with this volume and, for no apparent reason, provides the answer as 4 cm.

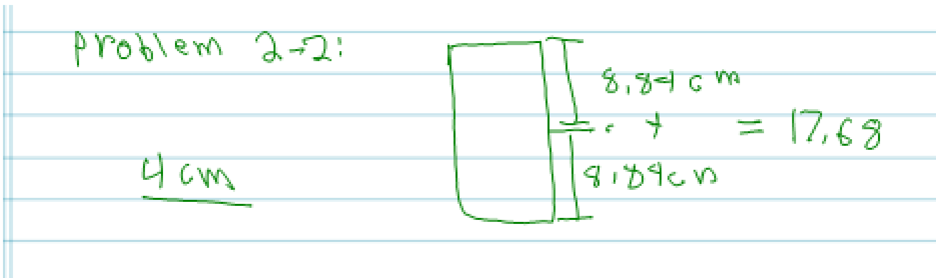


Figure 7-6. Jug problem solution from P22 (PSVT:R = 24).

Arguably, the most straightforward approach is to equate h = 8.84 cm with V = 1000 cm^3, Match 1. Both values are connected in the second sentence of the problem statement and can be directly matched without any arithmetic manipulation. This match requires a linguistic ability to ignore the reference to V = 2 litres in the first sentence – it is not relevant – and focus instead on the information in the second sentence. As shown in Figure 7-7, P6 notes the full volume of 2000 cm^3 but rejects its use in the solution and instead matches 1000 cm^3 with 8.84 cm. The match is made and included in the writing of the equation which is then rearranged to solve for r.

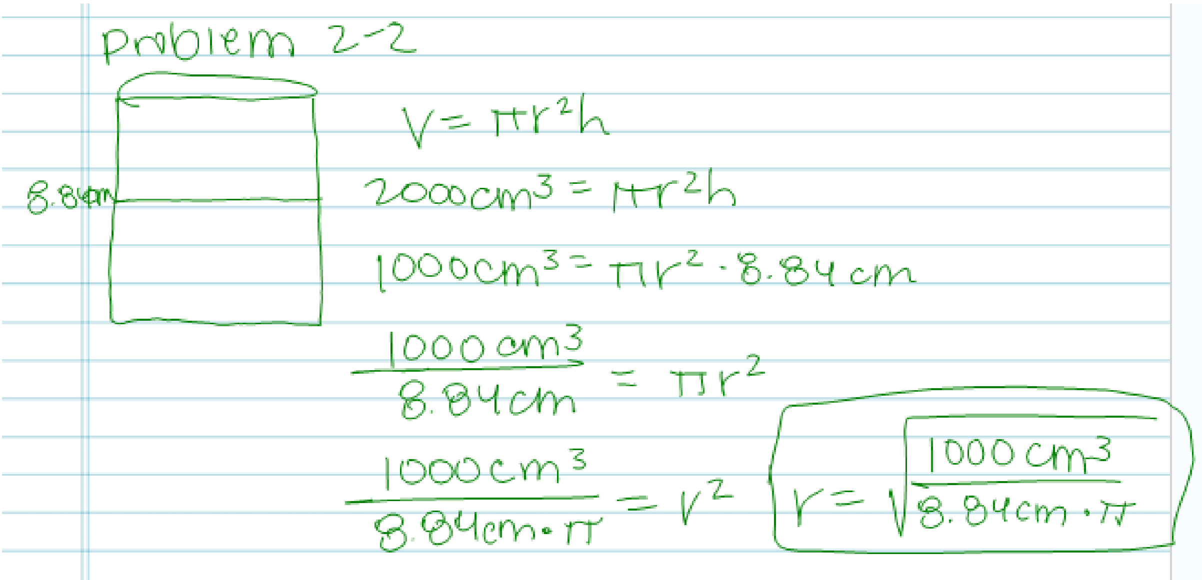


Figure 7-7. Jug problem solution from P6 (PSVT:R = 18).

It is also appropriate when solving this problem to match the full height with the full volume – h = 17.68 cm and V = 2000 cm^3 (or 2 litres) – this will also lead to the correct answer and was coded as 'Match2'. However, unlike Match1, the 35 participants (15 weak and 20 strong) who took this approach were no different from the rest of the sample in spatial ability as it was an approach equally favoured by both weak and strong visualizers. As shown in the solution of P26 in Figure 7-8, greater effort is needed to make Match 2 compared with Match 1 as the height provided at 1 litre, 8.84 cm, must be doubled to give 17.68 cm which is the corresponding height to 2 litres. Otherwise, the solution is the same as for Match 1.

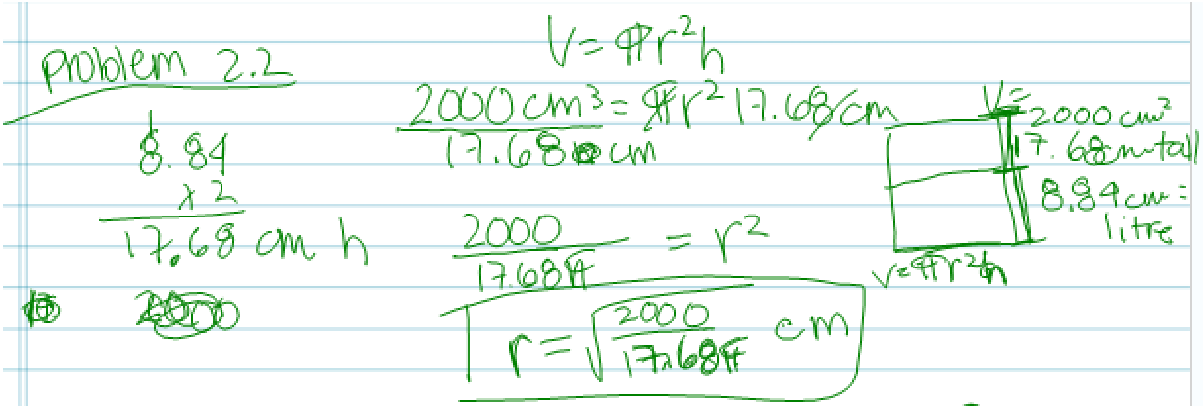


Figure 7-8. Jug problem solution from P26 (PSVT:R = 16).

Fifteen participants made errors in matching a height and volume, i.e. they provided a match other than Match 1 or 2. For example, in Figure 7-9, P38 matches h = 8.84 cm with V = 2000 cm^3 and also uses an incorrect expression for cylinder volume.

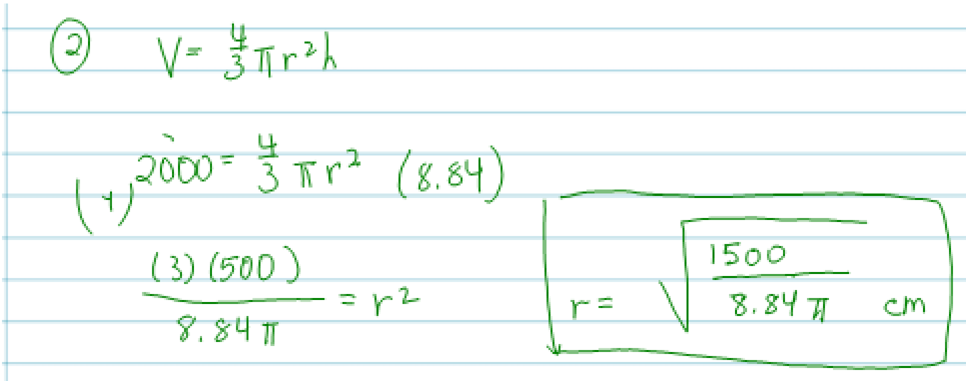


Figure 7-9. Jug problem solution from P38 (PSVT:R = 12).

The solutions provided by P6, P26 and P38 are qualitatively similar as they contain a cylinder volume schema, an expression for cylinder volume and a match between height and volume. The first two are error free while the third, P38, contains two errors, either of which leads to an incorrect solution. Although P38 has the right approach, differences exist at linguistic and semantic levels which result in the loss of points were this to be assessed in a manner typical for STEM education.

Summary of Jug Problem representation

The solution phase in this problem begins by writing an equation that connects the two products of representation – expression for cylinder volume and a match between height and volume – and concludes by rearranging this equation to solve for radius. However, as shown in the examples, it was not always possible to separately code the representation and the equation but it was possible to distinguish the next step of rearranging of the equation to provide the solution. All but 15, i.e. 100 participants, equated a volume and an expression for cylinder volume while 77 had both a correct expression and a correct match. Of these 77 participants, 69 correctly solved the problem.

There are two viable alternatives to characterising the key representation code for this problem. One is to consider the cylinder volume schema as universally adopted (all but 3 show evidence of it) and focus on the other ingredient needed, Match 1 or 2. Hence, one can conclude the representation of this problem hinges on the linguistic act of matching a height and volume appropriately. However, not all members of this group could solve the problem as some made errors in recalling the correct expression for cylinder volume ($\pi r^2 h$). This suggests an alternative grouping made from the combination of Match 1 or 2 and correct volume expression. While this divides the sample more evenly it straddles both the representation and solution phases of the problem. It was therefore concluded that Match 1 or 2 is the more appropriate code to select as capturing an essential feature of representation of the Jug Problem.

The Cans Problem

> *Drink cans are made by stamping out circular discs from a sheet of metal. The rectangular sheet from which the discs are stamped out measures 1 m by 2 m. If the cans have a radius of 10 cm, how many discs can be made from this sheet of metal?*

The Cans problem also contained a small number of steps to be completed. The metal sheet is clearly described in the statement as a rectangle measuring 1 x 2 m and this was correctly understood by all but two participants. There was variation in how this sheet should be manipulated to produce the circular discs with 8 distinct approaches evident in the data which could be grouped into three schemas – grid, zero waste and irrational. In the grid schema the sheet is divided into rows and columns whose widths are equal to the can top diameter. Zero waste is achieved by dividing the sheet area by the can top area. An example of an irrational schema is to divide an area by a length. Only one schema was appropriate – the grid schema – and within this, only one grid layout was correct – the 5 x 10 grid. The solution phase was manifest in moving from the schema of a grid layout to the particular 5 x 10 layout. The only calculations needed to make this movement were to (i) double the radius to give the diameter, (ii) divide the sheet width and length by the diameter to give 5 rows and 10 columns and (iii) multiply the numbers of rows and columns to get 50 discs, the answer. Strong visualizers were more likely than weak visualizers to select a grid schema, and of those who did, strong visualizers were more likely to divide the sheet into a 5 x 10 grid. The above data are summarised in Table 7-3.

		Evidenced by (%)			
Action	**Description**	**All**	**Weak**	**Strong**	**Cohen's d**
1	Read the problem	100	100	100	0
2	Select any grid schema	59	40	71	.81**
3	Select a 5 x 10 grid schema	39	22	55	.78**
4	Provide the correct answer	39	22	55	.78**
	(Select a zero waste schema		36	15	.59**
	Select an irrational schema		18	5	.51)

** Significant at the p < .01 level

Table 7-3. Actions evidenced by weak and strong visualizers on the Cans problem.

The following issues related to problem representation were observed in the Cans problem:

1. Strong visualizers favoured a grid schema while weak visualizers favoured a zero waste schema,

2. Almost one third of those who selected the grid schema made errors in its implementation and
3. A small number of participants (n = 13), mostly weak visualizers, selected an irrational method of calculating the number of can discs.

Since the recycling of material in a metal sheet stamping operation is a realistic option – a common goal in manufacturing is to minimize waste – it is reasonable that a zero waste schema would be chosen in solving this problem. On the other hand, since there is no mention of recycling in the problem statement this approach is not justified in this context. The most plausible and coherent interpretation of the problem statement is that waste is not recycled and the sheet should be divided into a grid with the row heights and column widths both equal to the diameter of the disc. Hence, the grid schema is the more appropriate. Within this schema there were two types of grid, straight and staggered rows, both of which are appropriate to the problem with the latter being a more advanced approach as it leads to less waste than the former. Staggering the rows was only attempted by two participants, both of whom provided an incorrect answer for this method, and they were treated as outliers and not included in the 5 x 10 grid group. The third schema to be used by a small number of participants, made no sense as it divided an area by a length which does not yield a quantity of discs. The area was that of the sheet and the length dimensions used included radius, diameter and circumference, all properties of the circle. Division of sheet area suggested a zero waste schema and this group could arguably be placed within that schema rather than separately. However, dividing by a linear property of the circle appeared to be a qualitatively different approach that did not fully embrace the zero waste concept. The sample was, therefore, grouped into three schemata – grid, zero waste and irrational – with variations in execution of the grid and irrational schemata. Two of these, grid and zero waste, were the predominant schemata.

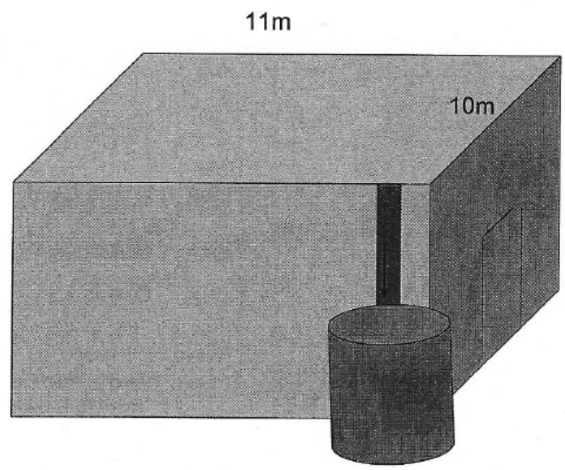
Strong visualizers tended to adopt a grid schema whereas weak visualizers tended to adopt one of the other two, zero waste and irrational. It appears that the zero waste schema reflects a lack of thinking about the meaning of the problem; the participant spots important variables in the problem statement and sets about manipulating them with little thought as to the context. The grid approach, on the other hand, reflects some consideration of the problem scenario, some thinking about context and what is happening in the stamping process. This leads to a consideration of the process from the perspective of the sheet and how it is transformed during the stamping. This reveals that when circles are removed some waste is left behind – the square bounding the circle with the circle removed. Since the problem does not state that this perforated sheet is recast as a new sheet that is complete but smaller the participant adopts the grid schema in which waste is produced.

Summary of Cans Problem representation

It was concluded that the key aspect of problem representation in this case was to apply 'any grid' layout, 5 x 10 or otherwise, to the metal sheet as this code split the sample somewhat evenly and revealed a large effect size in spatial ability. The nature of the representation challenge in this problem was schematic and not linguistic or semantic.

The Rain Problem

"*The diagram shows the dimensions of a flat roofed commercial shed. During one week 5 mm of rain fell on the roof of the shed. The rain was collected by gutters that flowed into a cylindrical water barrel with a diameter of 1 m. By how much did the depth of the*



water in the barrel increase as a result

of this rain?"

This problem proved to be quite challenging. It contained several sequential steps and, compared to the Lawn problem, had a similarly low success rate but a higher effect size ($d = .83$, $p < .001$). A diagram was provided which lessened the burden at the linguistic level – the participant did not have to translate the roof dimensions from text to mental model as the diagram presented a visualization of the roof and barrel. What remained was to assign 5 mm as the rainfall height and 1 m as the barrel diameter. The overall schema was that of a volume transfer from a prism to a cylinder. This first required a decision to use a prism volume equation to calculate the volume of rain. The remainder of the problem was then very similar to the Jug and Blood problems – recall the cylinder volume expression and solve for the one unknown, in this case height, using the two known parameters, volume and diameter. A challenge in this problem was to include the roof in the solution by calculating a roof volume and equating this with barrel volume change. Greater attention was paid to the barrel with many weak visualizers ignoring the roof volume altogether and instead seeing the rainfall height as a property of the barrel. The essential actions needed to solve the Rain problem are shown in Table 7-4 along with the numbers of weak and strong visualizers who showed evidence of each action.

Action	Description	Evidenced by (%) All	Weak	Strong	Cohen's d
1	Read the problem	100	100	100	
2a	Roof volume equation	62	46	77	.83**
2b	Convert units correctly	61	57	65	.51**
2	Calculate correct roof volume	47	34	60	.75**
3	Adopt volume is conserved schema	58	41	73	.89**
4	Adopt cylinder volume schema	92	89	95	.57 N.S.
5	Use correct cylinder volume equation	71	54	85	.97**
6	Solve for correct height	28	9	45	1.01**
	(Height conserved schema		34	13)	

** Significant at the p < .01 level

Table 7-4. Actions from the Rain problem.

Observations related to problem representation that were made from the Rain problem include:

1. Strong visualizers were more likely than weak visualizers to adopt the roof volume schema when solving this problem,
2. Strong visualizers were more likely than weak visualizers to treat the volume as being transferred from the roof to the barrel, i.e. from one container to another,
3. Weak visualizers were more likely than strong visualizers to see rainfall height/level as being conserved, and
4. Strong visualizers were more adept at handling units and recalling the cylinder volume equation and this is discussed in Chapter 5.

The first key step in this problem was to calculate the rain fall volume but many failed to see the need to do this as, in its appearing to them, the roof was not relevant to the problem and the problem statement did not evoke in them an image of rain falling on the roof which collected it for transfer to the barrel. This is illustrated in the transcript from P42 who begins the solution by redrawing the roof diagram and then says:

- P42: "*I'm a bit confused at the moment, I understand why it's relevant, the 5 mm, but I don't understand the relevance of the parameters given for the roof [P42 is referring to* *the dimensions, 10 and 11 m]. I'm thinking I should find the total area of the cylinder …. because it [the problem statement] gave me the diameter*".

At no point in the solution does P42 calculate the roof volume, it was completely ignored, but P42 did spend time calculating the cross sectional area of the barrel.

- P42: "*I understand the question but I'm confused by the information given … so the rain is collected by gutters that flows into a cylindrical water barrel with a diameter of 1 m … so would it not just be 5 mm? Because like, err…, no it would not be*."

P42 thinks a bit more but remains fixated on the barrel depth or height being equal to 5mm and ends the problem by providing 5 mm as the answer to the problem, albeit with some uncertainty.

Of the 27 participants who viewed the problem in this way, 19 were weak visualizers and 8 were strong visualizers leading to an effect size of .57, $p < .05$. The majority, 23, came from the OSU sample prompting the idea that the confusion might be cultural. However, since 30 of the OSU participants did not make this error and the majority of OSU participants were weak visualizers, the cultural explanation arguably accounts less for the difference than spatial ability.

Another example of this confusion is illustrated in the excerpt below from P21:

- P21: "*So 5 mm fell across the entire roof, I guess I don't understand the question*"
- Interviewer (I): "*It rains and 5 mm of rain fell on the roof*"
- P21: "*Like the height of it was 5 mm across the entire roof*?"
- I: "*Yes*"
- P21: "*OK*", said in a happy way and then proceeds with problem solving.

The role of the interviewer as observer was challenged here as the intention was to neither help nor hinder the participant. In response to the failure to understand the question, the part the problem statement which says “*5 mm of rain fell on the roof*” was repeated rather than providing another a reworded version of the problem statement. Rather than confirming P21’s view of rainfall height on the roof as 5 mm, perhaps the response should have been ‘*What do you think?*’ instead of ‘*Yes*’ but P21 had initiated this representation independently. P21 then proceeded to calculate the correct volume, shown in Figure 7-10, but failed to equate this with the cylinder volume expression and, hence, failed to solve the problem.

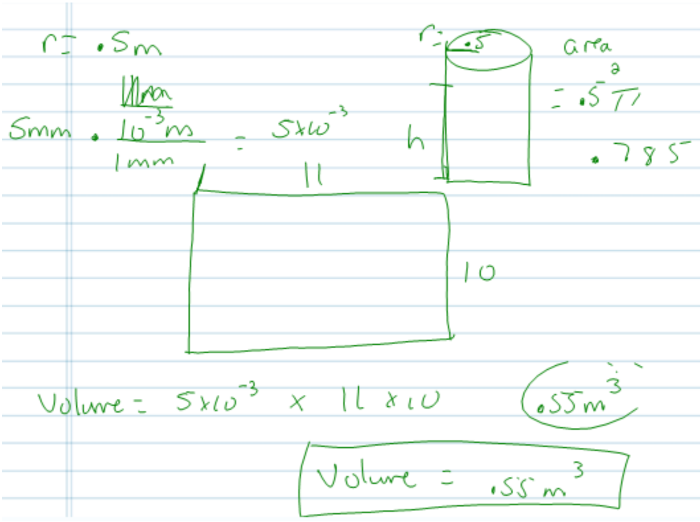


Figure 7-10. Rain problem solution from P21 (PSVT:R = 17).

P19 began to solve the problem by writing the correct equation for the volume of the cylinder ($V = \pi 0.5^2 h$) but then described the challenge of representing the problem:

- P19: “I’m trying to think of how to use the measurements of the shed to find the volume of the cylinder. It looks like it’s half of the height [P19 is looking at the diagram] but I don’t know the height of the shed either” [It’s not provided and not required]
- P19: “I can find the area of the roof but I don’t know how that would help me”.

- P19 then proceeds to do this calculation (11 x 10 = 110 m^2) but is still searching before saying "*Please tell me this is a trick question*".

P19's solution is provided in Figure 7-11

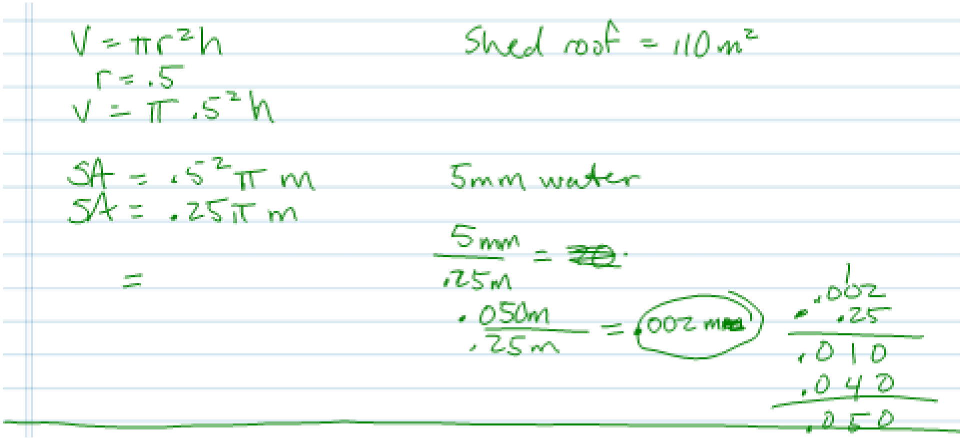


Figure 7-11. Rain problem solution from P19 (PSVT:R = 24).

P11 is aware of the role of the gutters but describes a schema that is not consistent with volume transfer and conservation:

- I: "*What is going on in this problem*?"
- P11: "*If the rain is being collected into the barrel, and there are gutters on each side then all of the water would be collected into the barrel, so if only 5 mm of water fell then that's how much would be going into the barrel. I think that's it. I think that's what it's asking me*". [P11 provides the answer as 0.005 m]

Another interesting point about this transcript is that little attention was paid to the roof, its dimensions, its purpose and the possibility that water could pool on the roof so that it is like a tank. The barrel attracted the attention. This was a common theme for many participants who failed to see volume as being conserved. P52 stated "*it seems that it doesn't really matter the dimensions of the roof if all the water went down* [sic]".

Figure 7-12 shows the entire solution from P39 and is another example of ignoring the roof entirely and focusing on the barrel. Even though several lines of calculation are provided here, the final answer, were it to be computed, would have been 5 mm x πr^2, i.e. a barrel volume based on the height of rainfall provided in the question. The roof is nowhere to be found in the solution.

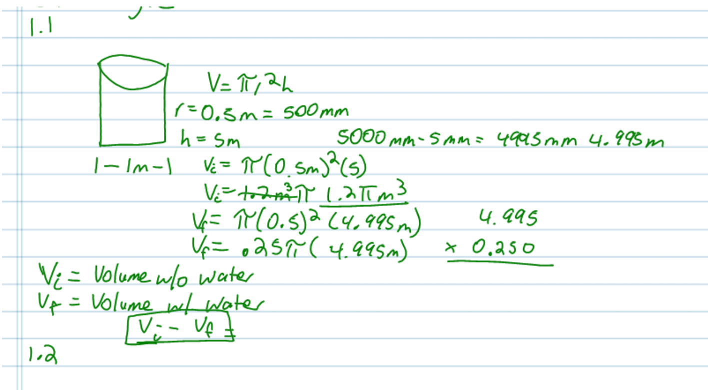


Figure 7-12. Rain problem solution from P39 (PSVT:R = 16).

Many participants had little difficulty in getting started as illustrated by P17:

- P17: "*Find the area first, 10 x 11 so it's a 110 m², and it's eh 5 mm deep …. so it's 0.550 m³*"

P17 then equated this with cylinder volume and solved for h to get the correct answer.

Another example of a correct solution is provided in Figure 7-13 from P46 who quickly adopted the schema of volume conserved and transferred. Having read the problem out loud, P46 immediately commented "this is basically a volume problem". Interestingly, attention was first focused on the barrel and its volume but, having written the cylinder volume expression, P46 identified h as the unknown in the expression as the target of the problem solution. The schema is fully developed on paper by line 3 and P46 replaces h of the cylinder with d to distinguish it from h of the roof. P46 then proceeds to solve the problem with ease.

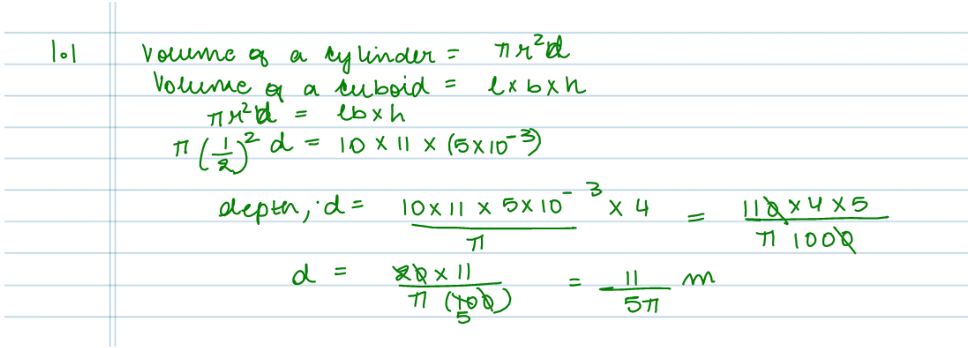


Figure 7-13. Rain problem solution from P46 (PSVT:R = 10).

Summary of Rain Problem representation

In summary, participants were challenged at semantic and schematic levels in this problem. The problem stated the barrel was a cylinder and the majority of participants showed some evidence of this in their solutions. Providing the correct cylinder volume expression was challenging and split the sample by spatial ability. Many failed to represent the roof as a prism whose height is rainfall with most of these being weak visualizers. The related schema of volume transferred and conserved also split the sample into large portions and to a significant extent by spatial ability and appears to be an essential aspect of representing the Rain problem.

The Pencils & Jars Problem

I have some pencils and some jars. If I put 4 pencils into each jar I will have one jar left over. If I put 3 pencils into each jar I will have one pencil left over.

How many pencils and how many jars are there?

At the research design phase it was assumed that participants would attempt to solve the Pencils & Jars problem algebraically by writing two simultaneous equations and solving for the numbers of pencils and jars. In this case, the problem representation phase would be manifest in the creation of the equations and the solution phase in their manipulation to extract the unknowns. Hence, the inclusion of a core competency question to separately test the solution phase. This assumption ignored the creativity and individuality expressed by participants when problem solving. Many chose to guess and check and this approach proved to be a viable and alternative pathway from problem statement to solution. Therefore, the first choice was not how to represent the problem algebraically but whether to use algebra at all or to guess and check instead. This may not have been a discussion at all. Some may simply have immediately seen an algebraic solution as the only option while others assumed guess and check to be the only option. Several participants tried both methods indicating that for them, a choice existed. This presents a new theme in the data analysis as the data effectively appear to come from a research design in which some participants had been told to solve this problem by guess and check while others were told to solve through algebra.

Since each approach has different steps and requires different competencies it makes some sense to group the responses based on approach and analyse them separately. Steps are easily defined for the algebraic approach – translate each relational statement into an equation, manipulate the equations and solve for the two unknowns. However, the guess and check approach, by definition, contains two interdependent steps – one guesses, then checks and continues this loop until the answer is obtained. The steps come as a pair and never alone – one does not guess without check or check without guess. One can pick any number as a starting point and proceed from there by increasing or decreasing if the guess is wrong. Guess and check requires the same rules as the algebraic method, the difference being that the rules are kept as relational statements rather than algebraic equations. Participants who adopted the algebraic method could be scored for each step whereas those who used guess and check only be scored on this single code. The problem solving actions for the Pencils & Jars problems are provided in Table 7-5 along with the numbers of weak and strong visualizers who showed evidence of each.

Action	Description	Evidenced by (%) Weak	Strong	Evidenced by (n) Weak	Strong	Cohen's d
1	Read the problem	100	100	56	60	-
2a	Guess & Check approach	84	50	47	30	.89**
2b	Algebraic approach	38	55	21	33	.50**
3a	1 correct equation	14[1]	6[1]	3	2	-
3b	2 correct equations	0[1]	12[1]	0	4	-
3	Provide the correct answer	46	30	26	18	.15

[1] as a percentage of those who took the algebraic approach

Table 7-5. Key actions on the Pencils and Jars problem.

Observations related to problem representation include:

1. Weak visualizers favoured a guess and check approach while strong visualizers favoured the algebraic approach,
2. Translating relational statements to algebraic equations was very challenging and the few who successfully modelled the system had very high spatial test scores.

Guess and check typically proceeded as follows: pick a number of jars and draw them, then fill each jar with 4 pencils by drawing 4 lines in each jar, then draw one extra empty jar; now check this against the second statement by drawing another set of jars with 3 pencils in each; if the two sets match the problem is solved, if not, then either a jar is added or removed and the process repeated. This is illustrated in Figure 7-14 in which P37 began by drawing 2 jars, filling each with 4 pencils, and then drawing one empty jar – this satisfied the first statement. P37 then checked these numbers against the second statement by drawing a second row of jars and filling each with 3 pencils and drawing one extra pencil. Confusion could occur at this point. For example, P37 drew just 2 jars on the second row when it should have been 3. P37 paused for a while, read the problem again, then added an extra jar to the top row and two extra jars to the bottom row. Now P37 had corrected the first error by having the same number of jars on each row. The number of pencils was counted and found to not match up. So P37 filled the fourth jar on the top row with four pencils and drew one extra empty jar. The second row was then amended to have five jars with 3 pencils in each and one extra pencil. Now P37 counted pencils and found they matched up on both rows and the problem was solved. At no point did P37 mention the use of algebra as a way of solving the problem. At the beginning, while drawing the first jar, P37 stated "*I'm more of a visual person, if I see it, I can think of it better*".

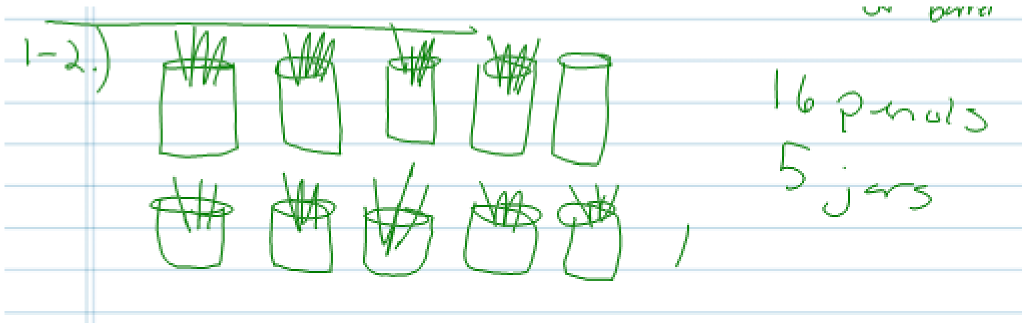


Figure 7-14. Solution to Pencils & Jars problem from P37 (PSVT:R = 17).

Rather than append extra jars some participants began a new row of jars for each new guess as shown in the solution from P43 in Figure 7-15

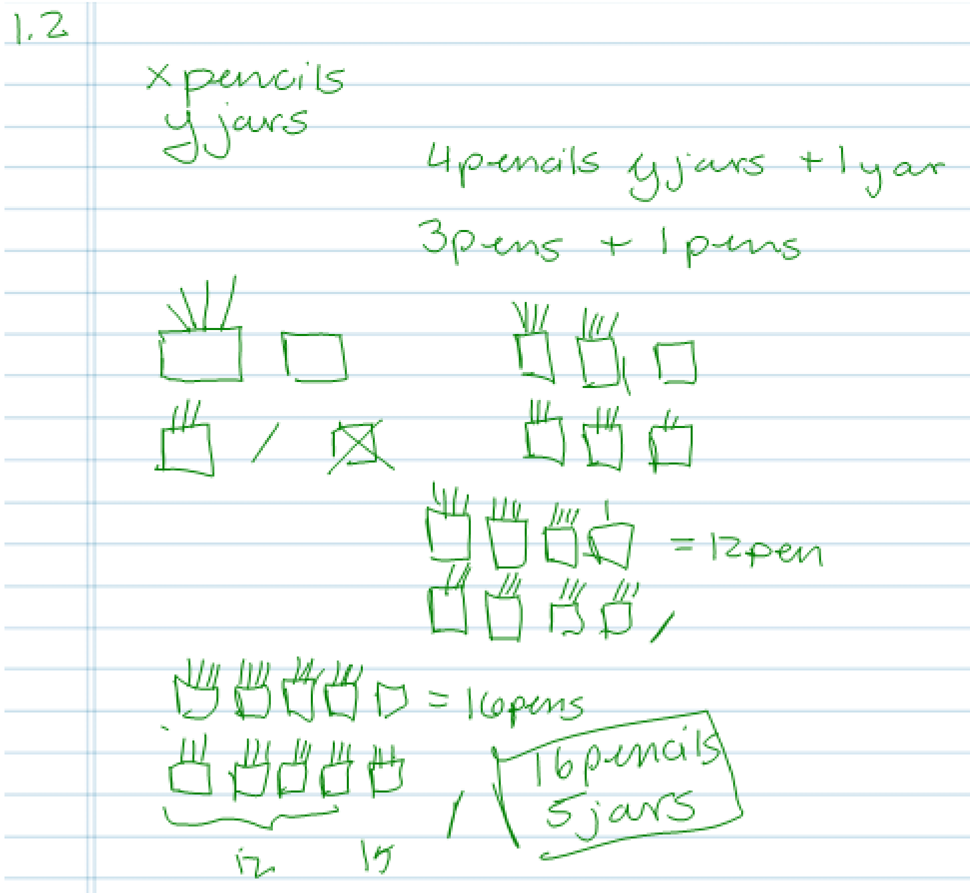


Figure 7-15. Solution to Pencils & Jars problem from P43 (PSVT:R = 15).

The solution from P38 (Figure 7-16) illustrates a very simple form of sketching with jars represented as lines that underscore a number of pencils represented as a numeral rather than lines. In this example, P38 has correctly understood the first statement but not the second.

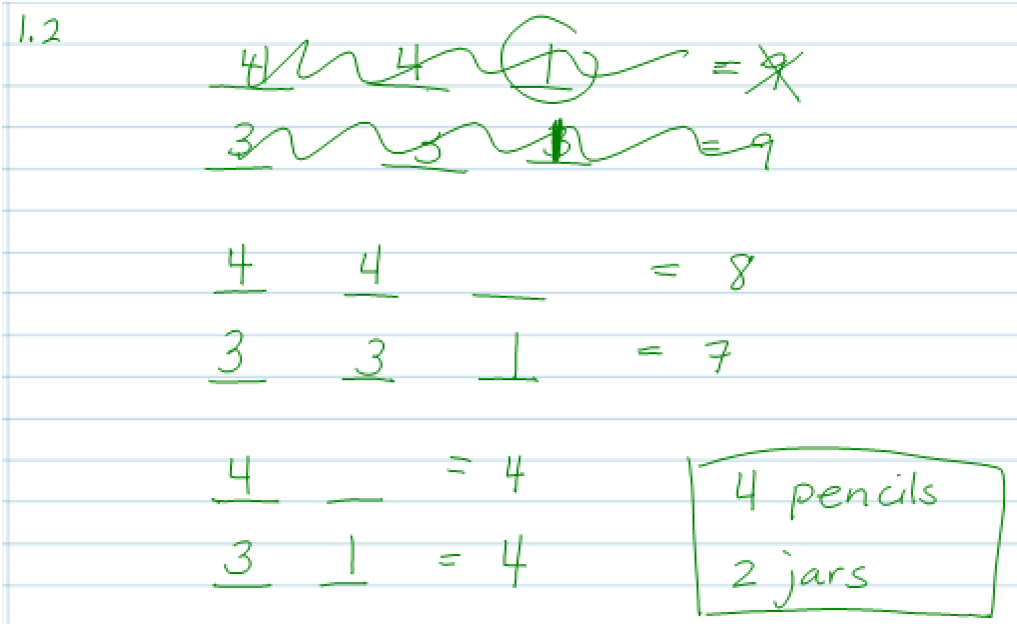


Figure 7-16. Solution to Pencils & Jars problem from P38 (PSVT:R = 12).

The provision of a sketch accompanied the guess and check approach in nearly all cases. Of the 77 who took a guess and check approach, 60 used a sketch of some form. Of those who took an algebraic approach, only 2 used a sketch. For many, sketching facilitated the guess and check approach – a participant typically guessed a number of pencils and/or jars and used a sketch to check the results of that guess. In ‘guess and check’ the sketch is the ‘check’. To check the results of the guess without drawing a sketch or writing a set of numbers is likely to place a larger demand on working memory. Sketching reduces that demand thereby allowing the guess and check method to be effective. Modelling, on the other hand, did not require sketches of pencils and jars. Here, one wrote equations and then checked them against the problem statements. Since strong visualizers were more likely to take this approach, they were also less likely to use a sketch. It is inappropriate, I believe, to interpret the results as indicating that sketching increases the success rate. Rather, a guess and check approach in which sketching acted as the check was an effective way to solve this problem.

A non-pictorial version of guess and check was used by 17 participants in which jars and pencils were not drawn but the numbers of each were written down. A variation on this alternative method of solving the problem was to realise that the number of pencils must be an integer, say n, that has 4 as a factor (rule 1) and that n-1 has 3 as a factor (rule 2). Instead of sketching jars and pencils one thinks of the multiples of 4 and examines them for compliance with rule 2. This was also a guess and check procedure which was evident in n=14 responses.

Of the 77 participants who used guess and check, 19 of these also attempted to model. Since 19 showed evidence of using two different schemata it is not possible to perform an independent samples t-test of those who did guess and check versus those who attempted to model. Instead, a comparison can be made between those who did guess and check versus those who did not (n=77 v 39) and also between those who attempted to model and those who did not (n=54 v 62). The number of weak and strong visualizers in each of these categories is shown in Table 7-6.

Visualizer	Total	Guess only	Model only	Guess & Model	No schema
Weak	56 (100 %)	33 (59 %)	7 (13 %)	14 (25 %)	2 (4 %)
Strong	60 (100 %)	23 (38 %)	26 (43 %)	7 (12 %)	4 (7 %)

Table 7-6. Number of participants in each visualization category that adopted each schema.

The solution from P60 shown in Figure 7-17 is an example of a successful algebraic solution. Very few of the 54 who attempted to model were successful in doing so.

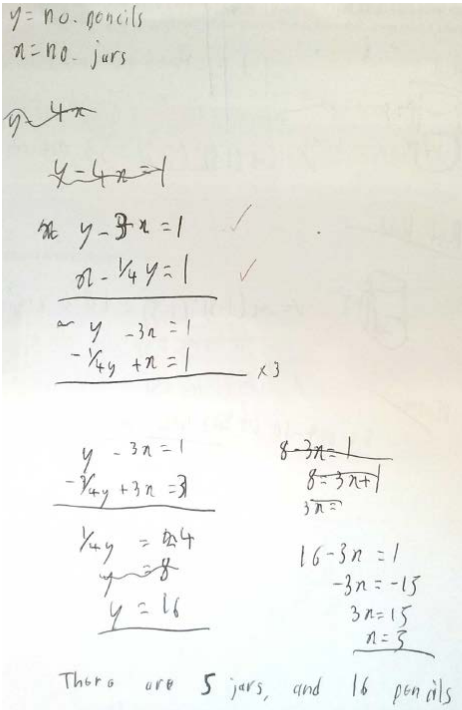


Figure 7-17. Solution to Pencils & Jars problem from P60 (PSVT:R = 28).

While significant spatial ability differences were revealed by both representation codes – guess and check and algebraic method – the former revealed the larger correlation as shown in Table 7-7 suggesting it should be selected as a key code for this problem. On the other hand, making a decision to use algebra to solve the problem represents a qualitatively different approach that reflects an engineering or mathematical mind-set. When given free choice, some choose to think mathematically while others do not and the algebraic solution method can be seen as distinguishing in this way. The absence of guess and check represents a decision to not use the less “engineering” approach (and to have no representation at all) whereas the algebraic approach is to act in a way that is positive from a STEM perspective. Therefore, the algebraic approach was selected as the key code for this problem.

PSVT:R score v	r^2	F	p
Guess & check	.143	18.515	.000
Model attempt	.065	7.754	.006
Model quality	.064	7.617	.007

Table 7-7. Regression of spatial ability with each representation option.

The results show that when allowed to solve this problem in a way of their own choosing, those who make the decision to use mathematical modelling have significantly higher spatial ability. This problem, in its appearing to the participant, either evokes a guess and check schema or a modelling schema in nearly all cases. Those for whom the problem evokes a modelling response, albeit with varying degrees of success, have higher spatial test scores than those who take a guess and check approach. This shows that strong visualizers are more likely to attempt the translation of word statements into mathematical equations. Some may argue that this is what engineers are expected to do. The engineering attitude is not manifest in repeated guessing and checking but in the application of mathematical skills to develop models and algorithms that lead to a solution.

However, that is not to say that the only valid way to represent this problem was through modelling as there were other successful approaches that could be followed. Mathematics were required to implement guesses and check them. Indeed, thinking about a number, n, which has 4 as a factor and when decremented by 1 (n-1) has 3 as a factor reveals insight that is mathematically based. In other words, there were several ways to represent this problem and while each representation required mathematics, their relationships to spatial ability were not consistent.

This problem was unique in the set of problems used in the study in that it revealed no significant difference in success rates between weak and strong visualizers. What it did reveal, however, was a difference in approach to solving the problem favoured by each group but this difference in approach did not lead to a difference in success rate. Weak visualizers favoured guess and check-by-sketching which turned out to be an effective method of solving the problem. Strong visualizers, on the other hand, favoured translating the statements into mathematical equations but few were successful in solving the problem this way. As a backup approach, guess and check was used to rescue the problem solution.

If modelling is the engineering method, then strong visualizers are more likely to demonstrate an engineering epistemology and attitude than weak visualizers. A problem such as this, in its appearing to the strong visualizer, is best solved through modelling. However, only a minority – 4 participants who happened to have very high spatial test scores – were sufficiently competent at modelling to solve the problem in this way.

Summary of Pencils & Jars Problem representation

The take home message from this problem is twofold. First, strong visualizers are more likely to model the system and weak visualizers are more likely to solve this type of problem through guess and check. Second, high levels of competence in modelling a system like this accompanies high levels of competence in the PSVT:R. The first statement can be statistically supported by the data collected in this study but the same cannot be said of the second since the number of participants was relatively small. Further data are needed to verify this statement which must remain as a hypotheses for now.

The Blood Problem

When blood samples are centrifuged the blood separates into two distinct layers – one made up mainly of plasma and the other made up of red blood cells.

A sample of blood was put in a flat bottomed test tube with a diameter of 3 cm. When the blood sample was added to the tube it filled the tube to a depth of 7.5 cm. After centrifuging, the red blood layer was 1.5 cm deep. What volume of plasma was in the sample?

The first sentence in the Blood problem explains how a liquid mixture is separated into two layers which happens in a container that is partly described in the second sentence. The dimensions of the two layers are provided but the shape of the container is not explicitly stated, that must be inferred from '*flat* bottomed *test tube with a diameter'* as being a cylinder. Making that inference is the representation phase of the problem: it is a cylinder volume problem with height and radius known which can be used to calculate the unknown volume. All that is now required is to correctly recall the expression for cylinder volume. These steps and their success rates are outlined in Table 7-8.

		Evidenced by (%)			
Action	**Description**	**All**	**Weak**	**Strong**	**Cohen's d**
1	Read the problem	100	100	100	
2	Cylinder volume schema	80	69	90	.99**
3	Write cylinder volume expression	63	47	78	.85**
4	Compute radius	76	62	88	.94**
5	Provide correct answer	54	35	72	.86**

** Significant at the $p < .01$ level

Table 7-8. Key actions in solving the Blood problem.

The following issues related to problem solving were observed:

1. Strong visualizers were more likely than weak visualizers to select a cylinder volume schema,
2. Strong visualizers were more likely than weak visualizers to provide the correct expression for cylinder volume and
3. Fewer participants wrote the correct expression for cylinder volume in this problem when compared to the Jug problem.

The cylinder volume schema was adopted by the majority (80 %) of participants but by fewer weak than strong visualizers. As mentioned, the cylindrical shape of the test tube was inferred from the terms 'flat bottomed' and 'diameter', the latter indicating a circular cross section. Just under one third of weak visualizers failed to represent the problem in this way compared to one tenth of the strong visualizers. Problem solution was not possible without this representation. The next step, writing the correct cylinder volume expression was an element of this representation (correctly writing this expression indicated both the correct representation and expression, incorrectly writing the expression indicated the correct representation but not expression).

Of the 23 that did not adopt the correct schema, 7 provided a blank response, 9 solved for a difference in height, 3 viewed it as an area calculation, 1 treated the test tube as a prism and 3 provided a response that could not be attributed to any schema (e.g. a guess). The most common error at the representation level was to treat the problem as a height difference calculation rather than a volume calculation. To treat it so was to ignore the diameter provided in the statement. It's not clear why some participants failed to include diameter and focused instead on height differences only but an example of such an approach is provided in Figure 7-18.

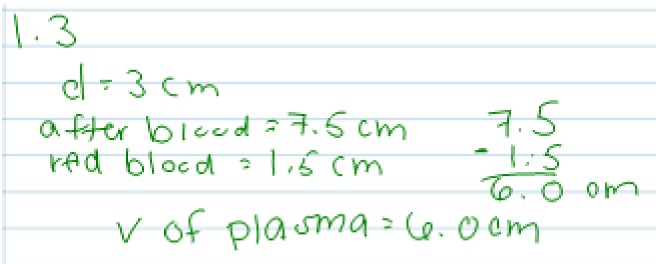


Figure 7-18. Solution to the Blood problem from P13 (PSVT:R = 16)

Despite writing down the diameter P43 ignored it in the problem solution (Figure 7-19). P43 was also happy to use units of cm for volume when it should be cm^3. At no point in the solution did P43 appear doubtful about the solution method. It appears reasonable to assume that this error is an oversight on the part of P43 who would agree that volume has units of length cubed if attention had been drawn to this error. However, there is no denying that this is how P43 represented this particular problem at a moment in time and failed in this case at the representation level by ignoring diameter and not treating the test tube as a cylinder. At this point in time, this problem appeared to P43 as one of a difference in heights which were labeled as volumes.

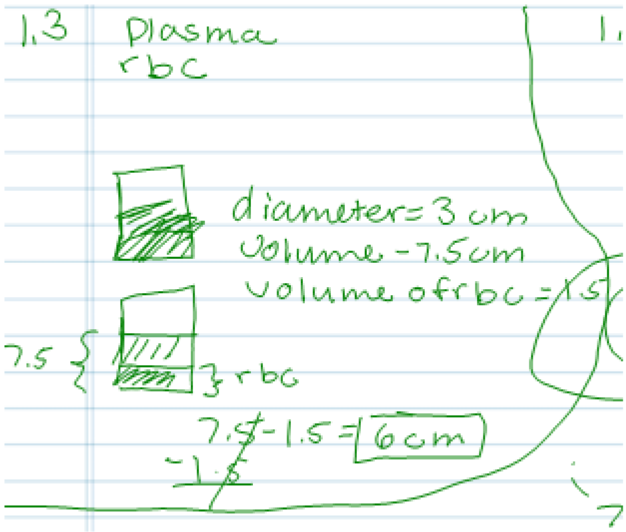


Figure 7-19. Solution to the Blood problem from P43 (PSVT:R = 15)

Summary of Blood Problem representation

Of the codes that were significant, the most important was the cylinder volume schema code as it contains all the others. While radius was needed to calculate the correct solution, all but two who calculated a radius also calculated a cylinder volume. Likewise for having the correct equation for cylinder volume. Deciding to calculate a plasma volume is a code made redundant by the 'Schema = Volume' code. Since 'Schema = Height' is its opposite, it too is included. Hence, the solutions to this problem can be divided into two groups – those who decided to calculate a cylinder volume and those who didn't.

The problem statement does not contain the word cylinder but that did not rule out the cylinder representation as one could easily arrive at this representation in two ways. Semantic knowledge that test tubes are cylinders is common and may be recalled. This is likely to bring with it the image of a test tube with hemispherical ends, hence the 'flat bottomed' adjective in the statement to simplify the geometry and calculations. The cylinder shape can also be inferred from the words diameter and volume that are included in the problem statement. Semantic knowledge should also lead to the image of a cylinder by connecting these two terms. The key issue in this problem was to visualize the test tube as a cylinder and solve accordingly. 90 % of strong visualizers took this approach compared with 69 % of weak visualizers.

Summary of problem representation and schema development

The purpose of this round of analysis was to identify a key aspect of representing each problem that split the sample somewhat evenly and revealed a large difference in spatial ability. As described by the Mayer schema, problem representation draws on linguistic, semantic and schematic knowledge. For two of the problems, Lawn and Jug, the key aspect of representation was linguistic in nature while the key aspect of representation for the other four problems was found to be schematic. At this stage, however, it was not clear if there was any pattern or consistency in representation among the participants across all the problems. Having identified these key representation aspects it was now possible to determine if, for example, weak visualizers were more likely than strong visualizers to consistently misrepresent different math story problems.

7.2 Comparison of representation across problems

More can be learned about the relationship between problem solving and spatial ability by looking across the problems and comparing the problem representations that came to the fore to each other. The objective of this section is to conduct this analysis and present the findings from it. It begins by highlighting the themes that emerged when the findings from each problem were compared to the others. Each of these observations is then described and elaborated by grouping common codes and comparing the sum of their scores to the PSVT:R scores. The section concludes with a discussion of what this analysis suggests about the phenomenon being studied.

Theme	Lawn	Jug	Cans	Rain	Jars	Blood
Aspect of representation	Linguistic	Linguistic	Schematic	Schematic	Schematic	Schematic
Key representation code	3 linguistic ingredients	Match 1 or 2	Grid layout	Volume is conserved	Model attempted	Cylinder volume
Errors in relational statements	Not all 3 linguistic codes obtained				Difficulties in writing algebraic equations	
Errors in core competency		Cylinder volume expression				Cylinder volume expression

Table 7-9. Problem similarities and differences based on analyzing the solutions to the problems.

The cross cutting themes that emerged from the individual analyses of the problems are listed in Table 7-9 and are described as follows:

1. For the Lawn and Jug problems the aspect of problem representation that challenged participants and revealed a difference in spatial ability was linguistic in nature. For these two problems, schematic representation was not challenging;
2. For the remaining problems, Cans, Rain, Pencils & Jars and Blood, the opposite was the case –schematic rather than linguistic representation came to the fore;
3. For the Lawn and Pencils & Jars problems, many participants failed to correctly translate all the relational statements in each problem highlighting a difficulty in being consistent at performing this task correctly;
4. For the Jug and Blood problems, many participants failed to solve these correctly because of errors in the cylinder volume expression and this revealed a difference in spatial ability;
5. The Pencils & Jars problem was the only problem not to reveal a difference in spatial ability between those correct and incorrect on this problem. The five remaining problems did reveal a difference in spatial ability.

Having identified these common themes, similarities and differences, the next step was to test the extent to which these themes correlate with spatial ability and determine how much of the relationship between spatial ability and problem solving they captured. This was done by creating a combined score from codes that are deemed to be similar so that if there are n codes each scored as 0 or 1, the range of scores for the combined code is 0 to n. This new measure can then be correlated with spatial ability. The sum of the corresponding problem scores is also calculated and correlated with spatial ability. If the correlation between the PSVT:R and this new combined grouping is at least as large as the correlation between the PSVT:R and the corresponding problem scores then this grouping of codes has captured an underlying pattern in problem representation that holds across problems and accounts for all of the variation between spatial ability and problem solving success. On the other hand, if the correlation of the PSVT:R with the group of codes is less than that with the problem scores then this aspect of representation partly explains the reason why spatial ability is correlated with success in the problems and other reasons also exist as to why success rates among weak and strong visualizers differ.

Linguistic representation in the Lawn and Jug problems

Both the Lawn and the Jug problem representations proved to be challenging at the linguistic level – it was difficult to discern all three linguistic aspects of the Lawn problem and only half the participants made the most direct match between height and volume in the Jug problem (Match 1). However, a majority made either Match1 or Match2 in the Jug problem, either of which could be used to solve the problem. This provides two options for combining or summing the linguistic codes from the Lawn and Jug problems – (1) All 3 linguistic ingredients + Match1 or (2) All 3 linguistic ingredients + Match1 or Match2. The correlations between the PSVT:R and each of these combinations is provided in Table 7-10. Included in the correlation matrix is the sum of scores for these two problems (range = 0 to 2).

	Sum of linguistic 1 (All 3 lnguistic ingredients + Match1)	Sum of linguistic 2 (All 3 linguistic ingredients + Match1 or 2)	Sum of problem scores (Lawn + Jug)
PSVT:R	.368**	.401**	.438**

** significant at p < .001

Table 7-10. Correlation matrix for PSVT:R, sum of linguistic codes and sum of problems for the Lawn and Jug problems, n = 115

In both combinations of linguistic codes the correlations with the PSVT:R were lower than the correlation between the PSVT:R and the combined scores on the two problems. This means more scatter and a less clearly defined relationship between the codes, which are steps towards the solution, as compared with the solution scores. More variation with the PSVT:R is shared with the second pairing compared to the first implying that Match1 or 2 is a code that captures more of the spatial/problem solving relationship. However, the correlation values are not that high implying that consistency in linguistic representation is not strongly related to spatial ability and suggesting that linguistic representation may not be a dominant feature of the relationship between spatial ability and problem solving skill as revealed by these problems. In other words, the linguistic operation as defined by either combination is an important aspect but does not explain in full the relationship between spatial ability and success in solving the Lawn and Jug problems.

Schematic representation in the Cans, Rain, Pencils & Jars and Blood problems

Schematic aspects revealed significant differences in spatial ability in the Cans, Rain, Pencil & Jars and Blood problems but not in the Lawn and Jug problems. For these latter two problems, the appropriate schemata, area and volume respectively, were identified by the vast majority of participants and success on these problems hinged more on the linguistic aspects of representation. The words ‘area’ and ‘volume’ were contained in the Lawn and Jug problem statements, respectively, and this may have allowed these schemata to be more easily cued. Equivalent words were not provided in the Cans, Rain and Pencils & Jars problems which meant the problem solver had to invest more cognitive effort in identifying or selecting them. The word ‘volume’ was contained in the Blood problem yet several participants failed to select a volume schema and this revealed a significant difference in spatial ability. Those who failed to identify the volume schema were in the minority (20 %) so although the effect size was large the sample was split very unevenly. In contrast, the success rates at the schematic level were much lower on the Cans (59 %), Rain (58 %) and Pencils & Jars (38 %) problems – for these problems, the effect size was large and the sample was split more evenly. Many participants were drawn towards alternative schemata for each of these three problems: zero waste for the Cans problems, rainfall = barrel height for the Rain problem and guess and check for the Pencils & Jars problem.

Leaving the Blood problem aside for the above reasons, the schematic code scores (0 or 1) for the remaining three problems were summed (Cans: Any grid, Rain: Volume is conserved and Pencils & Jars: Model attempted, range = 0 to 3) as was each participant's score on these three problems (range again = 0 to 3). Each of these measures was then correlated with the PSVT:R score with the results provided in Table 7-11.

	Sum of schema (any grid + volume conserved + model attempted	**Sum of problem scores (Cans + Rain + Pencils & Jars)**
PSVT:R	.533**	.331**
Sum of schema		.388**

** significant at p < .001

Table 7-11. Correlation matrix for PSVT:R, sum of schema and sum of problems for the Cans, Rain and Pencils & Jars problems, n = 115

In this case, a higher correlation was measured between the PSVT:R score and sum of schema than with the sum of the problem scores. More variation in the spatial test scores is shared with schema consistency (28 %) than with the sum of the three problem scores (11 %). To solve the problems correctly one must select an appropriate schema and do other things. Spatial ability shares more in common with the schema part than with schema plus other things. Strong visualizers are more adept than weak visualizers at schema development and their difference on this aspect of problem solving is more pronounced than for the entire problem solving process. Many were successful on the Pencils & Jars problem without attempting to model or use algebra; they could bypass this schema and still solve. Presumably this is why the correlation with the scores is lower than with the schematic codes.

Translating relational statements in the Lawn and Pencils & Jars problems

Relational statements were contained in the Lawn and the Pencils & Jars problems presenting an opportunity to compare scores on this code across these problems. In the case of the latter, however, attempts to translate the relational statements to equations were only made by those who attempted to solve the problem through algebra thereby limiting the data set. Those who took a guess and check approach had to comprehend the rules but did not have to translate the statements to equations. These participants bypassed this translation task so there is no data on their ability to do such a translation. While both of these problems presented a challenge to translate relational statements, a full set of data only exists for the Lawn problem thereby making it impossible to group the two data sets together and search for a pattern.

One finding that does emerge is that there is a difference between the ability to comprehend the statements and the ability to translate them into algebraic equations. Of the 77 who used guess and check, 41 solved the problem while only 4 of the 54 participants who attempted to model were successful in translating both relational statements. Assuming that comprehension of the statements is required to solve through guess and check, it appears that comprehending the statements is much easier than translating them into algebraic equations. This implies that although comprehension of relational statements can exist at a qualitative level this does not guarantee they can be expressed as quantitative relationships.

Errors in recalling cylinder volume expression in Jug and Blood problems

The correct cylinder volume expression was required in three problems – Jug, Rain and Blood. However, at least two things must happen for the correct cylinder volume expression to appear in a solution – select a cylinder volume schema and write the correct expression.

Consistency therefore depends on consistency of choosing the schema and consistency in writing the expression. The former has already been discussed and the latter was measured by core competency question 2 which was discussed in a previous chapter. Therefore, this common theme is not pursued here.

The odd man out

Of the six problems used in the study, the exception to the rule was the Pencils and Jars problem. This was the only problem that did not reveal a difference in spatial ability between those who were successful and unsuccessful in solving the problem. Likewise for its corresponding test of core competency, the ability to solve for two unknowns in two simultaneous equations did not reveal a difference in spatial ability. In any event, those who solved the problem, with the exception of four participants, did so through a guess and check approach that bypassed the use of this competency. The decision to take this approach over the more mathematically minded one of using algebra revealed a difference in spatial ability that was of a medium to large effect size. This raises these questions:

Why was the Pencils & Jars problem solved with equal success by weak and strong visualizers? What characteristic(s) did it possess that might explain this observation? What characteristic(s) did it not have that the other five problems possessed with regard to spatial ability?

The Pencils & Jars problem is different from the others in that properties of shapes, such as volume, area or dimensions, are not involved. It should be categorised as an algebra problem, thereby placing it somewhat in common with the Lawn problem, but, as shown by the data, algebra can be avoided and the problem solved through guess and check. It is perhaps this feature that marks it out from the others as guess and check rarely featured as a solution method on the other five problems whereas it was the dominant successful method in the

Pencils and Jars problem. Problem representation could be bypassed in this case and this could be why the problem was solved with equal success by weak and strong visualizers.

When the problem is solved without algebra its categorisation as an algebra problem no longer holds. Were it to be rephrased so that participants were required to use algebra the results would have been very different as the ability of all participants to translate both relational statements would have been measures. The data collected indicate that this ability would correlate with spatial ability but this assertion is tentative and more data would have to be collected for a test of significance to be performed. It is therefore believed that adding a requirement to translate the statements to equations would fundamentally change the characteristic of the problem and a correlation with spatial ability would emerge. What the problem lacked and that the other problems possessed was a requirement to create a representation before solving. When viewed as an algebra problem it possessed this characteristic and success correlated with spatial ability (a tentative assertion). When viewed as a guess and check challenge, it lacked this characteristic and success did not correlate with spatial ability.

In hindsight, it could be argued that the problem should have been rephrased to require the algebraic approach. Without doubt this would have yielded some interesting results. On the other hand, the phrasing that was used led to an important finding: when given free choice, strong visualizers are more likely than weak visualizers to treat this as an algebra problem and weak visualizers are more likely than strong visualizers to treat it as a guess and check challenge.

7.3 Consistency of participants in representing problems

Some participants were consistent in misrepresenting problems, some were inconsistent and other were consistently good at representing problems. This section seeks to measure how consistency varied with spatial ability. It also examines the nature of consistency in problem (mis)representation by searching for patterns in participants' representations. For example, we read in the Jug Problem that "*If the 1 litre mark is at 8.84 cm, what is the radius of the jug to the nearest centimetre?*" yet only half the participants used these numbers in their solutions. How did each half translate the other problem statements? Could their translation of the Jug Problem be used to predict how they translated another problem? If so, then this reflects something happening at a psychological level; if not then these are potentially random acts with no underlying pattern.

Each problem, indeed every problem apart from Pencils & Jars, contains a key statement or group of statements that must be 'seen' by the participant for the problem to be solved successfully. While the Pencils & Jars problem did contain a key code – model attempted – it could also be equally well solved without a schema (by using guess and check). This presented two options for an analysis of consistency in problem representation and comparison with consistency in problem solution among weak and strong visualizers – (i) include the full data set bearing in mind that spatial ability is unrelated to problem score for one of the six problems or (ii) reduce the data set by removing the Pencils & Jars problem to leave five problems for which spatial ability is related to both problem representation and solution. Both options are next examined beginning with the reduced data set with Pencils & Jars removed. Hence, for the remaining five problems the key representation codes were:

- Lawn Problem: discern the three key statements in the problem - 'Lawn is square', 'New dimensions (width + 2, length + 3)' and '$A_{new} = 2 \times A_{old}$'.
- Jug Problem: make an appropriate match between height and volume.
- Cans Problem: select a grid schema in solving the problem
- Rain Problem: see volume as being conserved as it is transferred from one container, the roof of the shed, to another, the barrel.

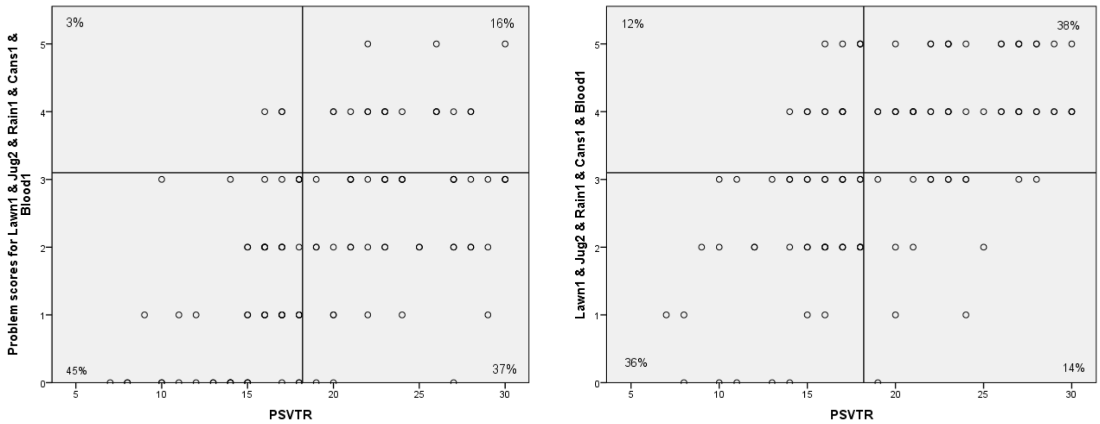
- Blood problem: represent the test tube as a cylinder whose property of interest is volume

		Weak n		Strong n		Total n		PSVT:R			
Problem	**Code**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**t (p)**	**Cohen's d**
Lawn	3 linguistic components	15	31	37	32	52	63	21.29 (5.29)	18.40 (5.58)	-2.833** (.005)	0.54 (Large)
Jug	Match1 or 2	32	14	60	9	92	23	20.71 (5.17)	15.70 (5.64)	-4.084** (.000)	0.93 (Large)
Cans	Grid layout	19	27	49	20	68	47	21.43 (5.30)	17.21 (5.14)	-4.242** (.000)	0.81 (Large)
Rain	Volume is conserved	16	26	51	16	67	42	21.57 (5.28)	17.48 (4.80)	-4.076** (.000)	0.82 (Large)
Pencil Jar	Model Attempted	15	30	37	31	52	61	21.25 (6.05)	18.33 (5.11)	-2.785 (.006)**	0.53 (Large)
Blood	Cylinder volume	31	15	61	8	92	23	20.75 (5.19)	15.52 (5.38)	-4.289** (.000)	0.99 (Large)

Table 7-12. A key code from each of the 6 problems.

In each of these cases, strong visualizers were more likely to adopt the successful approach or see the problem in an appropriate way, as shown in Table 7-12. However, these findings lack details in terms of the consistency in approach of weak and strong visualizers across problems. Not all of those who correctly interpreted the Rain Problem, for example, were successful in interpreting the other problems. Hence, it was decided to measure how the participants fared across problems as that would give an indication of consistency in approach to interpreting the problems. It was also decided to compare consistency at problem representation to consistency at problem solving in order to determine how much of the overall relationship between spatial ability and problem solving is captured by representation.

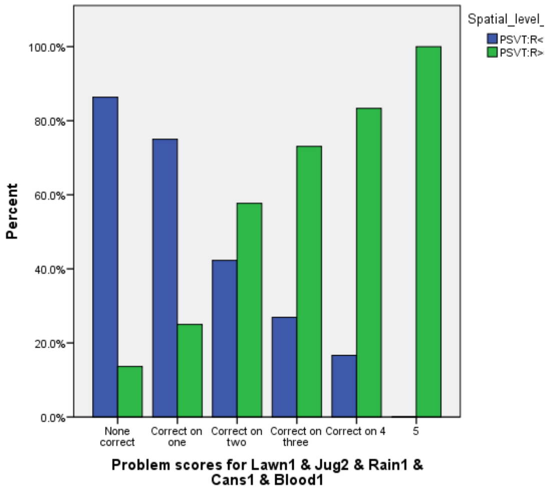
These results are presented first in Figure 7-20 as scatter plots of PSVT:R scores versus problem scores and problem representations. As these scatter plots reveal, the relationships between the PSVT:R and each of the two measures related to the five problems – problem scores and representations – are quite similar to each other which reflects a strong relationship between problem representation and solution, i.e. a correct representation is strongly predictive of correct solution and vice versa. However, when the scatter plot is divided into quadrants the distribution of participants does vary quite noticeably from one to the other with an upwards shift on the representation scores relative to the problem scores. The higher average scores on the representation codes than the problem scores occurs as participants can get the representation correct but subsequently make an error in solving the problem (e.g. recalling the wrong cylinder volume expression). There appears to be a more pronounced relationship between spatial ability and problem representation than between spatial ability and problem solution and this observation is examined by calculating a correlation coefficient for each relationship below.



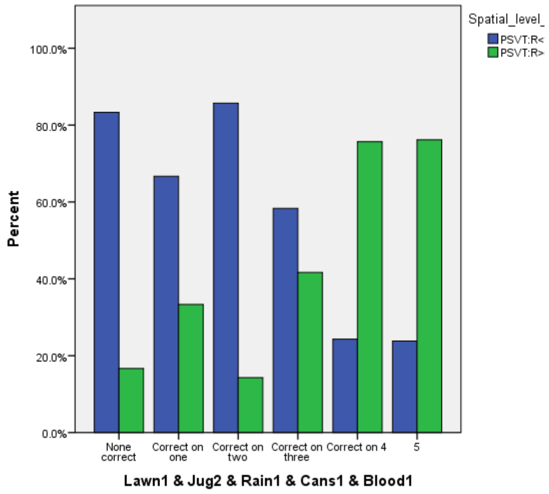
(a) (b)

Figure 7-20. Scatter plot of PSVT:R versus (a) problem scores (correct/incorrect) and (b) key representation codes for the Lawn, Jug, Rain, Cans and Blood problems (Pencils & Jars excluded).

Before examining correlations the data are next presented in clustered bar chart format to show the distribution of weak and strong visualizers when grouped by number of problems solved and by number of correct representations (Figure 7-21, with frequency on the y-axis expressed as a percentage of each x-axis category). The data used to create the bar charts are provided in Table 7-13 and Table 7-14 respectively. With regard to correctly solving the problems (Figure 7-21 (a) and Table 7-13), weak visualizers have a much lower rate of consistency when compared with strong visualizers with no weak visualizers present in the group that correctly answered all five problems.



(a)



(b)

Figure 7-21. Percentage by x axis category of weak and strong visualizers grouped by (a) number of correct problems and (b) number of correct representations for the Lawn, Jug, Rain, Cans and Blood problems.

Problem	Weak	Strong	Total	PSVT:R Mean	S.D.
Incorrect on all	19	3	22	14.18	4.584
Correct on one	15	5	20	17.35	4.428
Correct on two	11	15	26	20.62	4.622
Correct on three	7	19	26	22.46	5.255
Correct on four	3	15	18	22.72	3.739
Correct on five	0	3	3	26.00	4.000

Table 7-13. Distribution of weak and strong visualizers and PSVT:R scores grouped by number of problems solved correctly for the Lawn, Jug, Cans, Rain and Blood problems.

Membership of the groups is slightly different in terms of problem representation (Figure 7-21(b) and Table 7-14). For both weak and strong visualizers, more participants correctly represent than correctly represent and solve the problems which is to expected as the former is a subset of the latter but the data reveal a greater consistency among the strong visualizers.

Representation	Weak	Strong	Total	PSVT:R Mean	S.D.
Incorrect on all	5	1	6	12.50	3.834
Correct on one	4	2	6	15.00	6.633
Correct on two	18	3	21	16.19	3.628
Correct on three	14	10	24	18.50	4.836
Correct on four	9	28	37	22.30	4.858
Correct on five	5	16	21	23.43	4.296

Table 7-14. Distribution of weak and strong visualizers along with PSVT:R scores grouped by number of correct representations for Lawn, Jug, Cans, Rain and Blood problems.

A simple regression was used to measure the statistical properties of these relationships and these results are presented in Table 7-15 where it can be seen that spatial ability has a slightly stronger relationship with problem representation than with the combination of representation and solution. This implies that when grouped together, the codes that were identified as key aspects of representation do indeed capture an essential aspect of the problem solving process – the shared variation is no lower for the PSVT:R-representation code than for the PSVT:R-problem scores. In fact, the correlation with the representation codes is slightly higher with 1 % more variation shared with spatial ability.

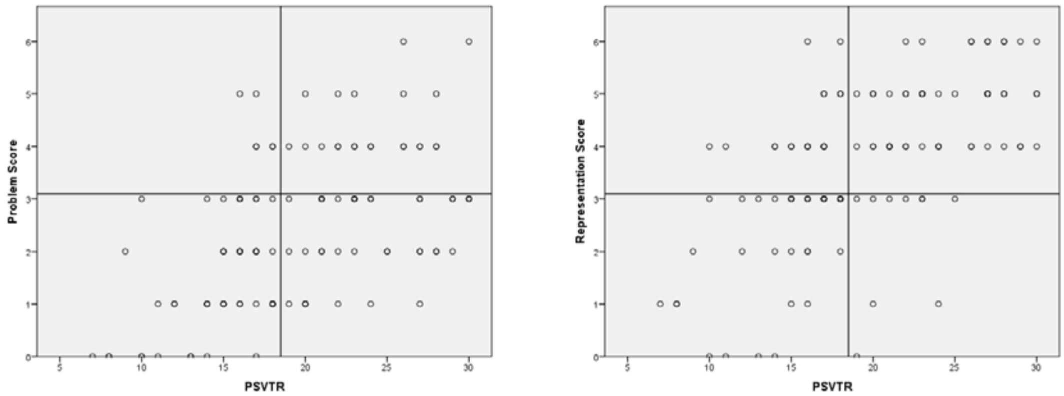
Regression of PSVT:R with	F(1, 114)	r (113)	r^2	p
Sum of key representation codes	58.919	.585	.343	.000
Sum of problem scores	56.299	.577	.333	.000

Table 7-15. Results from simple regression of PSVT:R with key codes and problem scores for the Lawn + Jug + Cans + Rain + Blood problems.

The scatter plots reveal another interesting aspect of these relationships. As shown in Figure 7-20 (a), the upper left region is mostly empty indicating that few weak visualizers are consistently successful at solving the problems. The upper right region, good problem solver/strong visualizer, is much more populated. In most cases, therefore, those with good problem solving skills also have good spatial skills and it is uncommon to find a good problem solver that does not have good spatial ability. The converse also appears to be true – those who have poor problem solving skills are also weak visualizers and these participants are found in the bottom left quadrant of the graph. The last quadrant, bottom right, is where those strong spatial ability and poor problem solving skills are placed and this region does contain a few participants. To conclude, one is most unlikely to be a good problem solver without strong spatial ability, a poor problem solver is likely to have weak spatial ability but some poor problem solvers do have strong spatial skills.

Figure 7-20 (b) presents the scatter of PSVT:R versus problem representation summed across the five problems. The overall appearance is similar to that of the problem scores but participants are grouped to a slightly greater extent in the upper right and lower left regions with noticeably fewer cases in the lower right region. The upper left quadrant is similar in each case. This indicates the trend established between PSVT:R and the problem scores is repeated but slightly more pronounced by the removal of some cases from the lower right, weak problem solving/strong spatial. Now what holds in the majority of cases is either good problem representation skills being accompanied by strong spatial skills or poor problem representation skills accompanied by weak spatial skills, i.e. a large majority of participants fall into one of these two quadrants on the PSVT:R versus problem representation scatter plot. In the next section, the approach of a participant from each of these quadrants is illustrated using examples from their transcripts.

The restriction of analysing all problems except Pencils & Jars was next removed and a similar analysis of consistency in representation and solution was conducted using the full data set. The key representation code for the Pencils & Jars problem that was added to the data set was 'model attempted', i.e. the decision to attempt to solve the problem using algebra. Presented in Figure 7-22 are the scatter plots of (a) the sum of representation codes and (b) sum of problems scores for all six problems versus spatial ability while in Table 7-16 are presented the Pearson correlation coefficients and regression values for these variables.



(a) (b)

Figure 7-22. Scatter plot of PSVT:R versus (a) problem scores (correct/incorrect) and (b) key representation codes for all six problems (none excluded).

Regression of PSVT:R with	F(1, 114)	r (113)	r^2	p
Sum of key representation codes	69.742	.618	.382	.000
Sum of problem scores	47.341	.543	.295	.000

Table 7-16. Results from simple regression of PSVT:R with key representation codes and problem scores for all six problems.

As shown in Figure 7-22, the findings are similar to the preceding analysis in which Pencils & Jars was excluded – data points are distributed in all but the upper left quadrant (low spatial, high math) for the problem scores while the distribution for problem representation is shifted with data points mostly found in the low spatial – low math and high spatial – high math quadrants. However, the difference is more pronounced for the set of six problems which is to be expected as performance on the Pencils & Jars problem, defined in this case by problem score, was unrelated to spatial ability and, therefore, less shared variation with problem score is to be expected. With regard to problem representation, however, the shared variation between spatial ability and the set of six problems (r^2 = .382) is higher than with the set of five (r^2 = .343) which reaffirms the existence of a consistent relationship between the two measures.

Poor problem solving skills/weak spatial ability

Six participants failed to correctly represent any of the five problems (Lawn, Jug, Cans, Rain and Blood). Of these, one had a PSVT:R score of 19 while the other five were weak visualizers.

P42 was one of these weak visualizers with a PSVT:R score of 8. For the first problem presented to her, P42 drew a promising sketch of the rain problem and wrote A x h = V but did not follow the volume trail and settled on barrel height as being 5 mm. P42's sketch of the test tube in the Blood problem looked like a cylinder with an elliptical top and base. Her plan was sound: find the total volume and subtract the red blood volume to get the plasma volume. Her error was to calculate volume as if it were area by multiplying the height by the diameter (7.5 x 3). She expressed doubt about this approach – "it doesn't seem quite right" - but finished the question by responding in a reasonably confident voice "*I believe so*" to her own check of the answer being correct. P42 provided the correct expression for the cylinder volume without difficulty when answering that core competency question. In the Jug problem P42 took the correct approach: she did treat the container as a cylinder by applying the correct cylinder volume expression. Her error was a simple one, she mismatched height and volume (h = 8.84 cm with V = 2000 cm^3) thereby getting the question wrong for linguistic and not schematic aspects of representation. One has to think that on another day P42 would make a correct match and that she is better described as someone who has a high likelihood of making this type of error on this type of problem rather than someone who is guaranteed to make these errors. She was occupied with an uncertainty about her volume calculation and perhaps this drew attention away from the matching process. She began to solve the Pencils & Jars problem by using a guess and check approach but abandoned this method after checking the first guess. She then switched to an attempt to translate the statements into algebraic equations. She persisted with this approach for a while and was successful in writing one of the two equations correctly but her failure to write the second resulted in her failing to solve this problem either through guess and check or with algebra.

Good problem solving skills/strong spatial ability

There were 21 participants who correctly represented all five problems (Lawn, Jug, Cans, Rain and Blood), 16 of whom were strong visualizers. P40 gained a PSVT:R score of 28 and correctly represented all five problems but correctly solved only two of these. He did not recall the correct cylinder volume expression due to mixing up area and circumference of a circle. P40 used $2\pi r$ for area and πr^2 for circumference of a circle and $2\pi rh$ for the volume of a cylinder when solving these problems. He made a very promising start to the Rain problem by describing qualitatively what was happening to the rain and how it was collected in the barrel. He then became slightly confused or distracted by wanting to know how tall the cylinder was, then wondering how much water was in the barrel to begin with. He then connected the 5 mm of rain directly with the barrel and appeared to be heading for rainfall = barrel height increase. At this stage, P40 was almost 3 minutes into solving the problem and had paid little attention to the roof of the shed. Like a number of other participants, he was almost exclusively focused on the barrel. Below is the section of the transcript at this point in the solution:

- P40: "*Wait, I'm over thinking this. Wouldn't it just be five millimeters? Oh but you want how much of the barrel. If five millimeters of rain is going to fall into the cylinder from the shed, if all that goes in, do we want to know by how much of the barrel distance the water would increase*?"
- I: "*It's how much the depth of the water in the barrel increased because of the rain*". [A paraphrase of the last line of the problem statement]
- P40: "*So we're assuming that the entire plane of the top of the roof had water on it*."
- I: "*The waterfall on the roof*."
- P40: "*Right*."
- I: "*And then it's collected and flows into the barrel*."
- P40: "*Right. So we didn't find the volume of the rain from the roof dimensions. So it would be eleven meters times two meters times five millimeters*."

He proceeded to solve the problem with the correct approach from this point forward but with the incorrect expression for cylinder volume.

In attempting the Blood problem P40 also uses the incorrect cylinder volume expression but does take the correct approach by subtracting red blood cell height from total height to give plasma height and then multiplying this by area to get plasma volume. His solution to the Lawn problem was direct, concise and error free as shown in Figure 7-23. Likewise for the Jug problem in which he made the direct match of height and volume (8.84 cm and 1000 cm^3) and solved the problem in a succinct manner. Finally, his solution to the Cans problem was also exemplary. He created a 5 x 10 grid based on converting the disc radius to a diameter and solved the problem correctly.

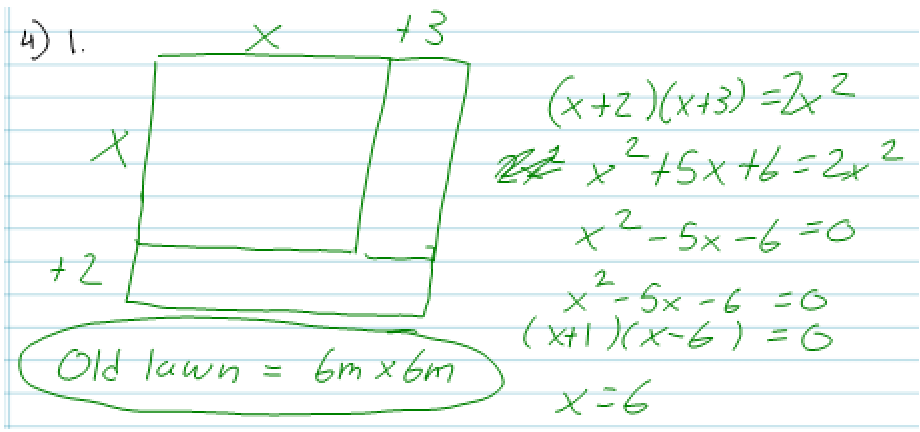


Figure 7-23. Solution to Lawn problem from P40 (PSVT:R = 28).

Summary and conclusions

A statistically significant relationship between problem representation and spatial ability was established in Chapter 5 which was followed in Chapter 6 by the development of a list of actions or steps required to represent and solve each problem as evident in the data. The purpose of this chapter was to reveal which of these steps were most strongly related to spatial ability. For each problem in turn, the problem representation codes were examined to identify key themes and, from these, select a key code that (i) captured an essential feature of problem representation, (ii) revealed a significant difference in spatial ability and (iii) divided the sample in two portions of similar size. Cross cutting themes that emerged included the dominance of linguistic aspects in problem representation for the Lawn and Jug problems and the dominance of schematic aspects of representation in the Cans, Rain, Pencils & Jars and Blood problems. When grouped together, the sum of the two linguistic codes had a slightly lower correlation with spatial ability than did the sum of the corresponding problem scores. When the schematic aspects were grouped, however, a higher correlation was measured. This suggests that a problem whose schema is easily identified but contains a challenge at the linguistic level is somewhat more likely to be solved by strong than weak visualizers whereas a problem whose schema is not self-evident is much more likely to be solved by strong than weak visualizers. It is suggested that this explains why the Pencils & Jars problem was solved with equal success by weak and strong visualizers - since a schema was not required to solve this problem weak visualizers were not disadvantaged. The gap between weak and strong visualizers is more pronounced when the representation challenge is schematic in nature.

When the key representation codes from each problem were summed the Pencils & Jars key code (model attempted) was excluded as it did not represent the only way to solve the problem. This allowed an examination of the relationship between consistency in correctly representing the remaining five problems and spatial ability which was compared to the relationship between spatial ability and consistency in solving these five problems. A slightly higher correlation was measured for the former relationship indicating that, at a statistical level, the PSVT:R scores shared slightly more variation in common with problem representation than with problem representation and solution. When the PSVT:R and problem solving data were plotted, cases were distributed across three quadrants – poor problem solving/weak spatial, poor problem solving/strong spatial and good problem solving/strong spatial. The scatter plot of PSVT:R versus representation revealed fewer cases in the lower right quadrant with the majority either poor problem representation/weak spatial or good problem representation/strong spatial. In this sample the ability to represent problems was accompanied by the ability to perform well on the PSVT:R.

This phase of the project began by administering three instruments – PSVT:R, a set of math problems and a set of math questions – to two samples of first year engineering students. These data were analysed in several ways. First to test the hypothesis provided in Chapter 5, second to qualitatively examine approaches to problem solving in the form of problem representation and solution steps and third to identify which of these steps were most strongly related to spatial ability. At this stage, all the outcomes that were identified in the research design had been completed. In the following chapter, the last in the thesis, a summary of all phases of the work is provided along with a critical analysis of each of the findings and what can be learned by comparing findings to each other.

Chapter 8 Discussion and conclusions

A motivation for this study was to contribute to the understanding of why spatial ability accompanies success in STEM education. High school students with the best scores on spatial tests enjoy taking STEM courses and are attracted to engineering education. Once there, strong visualizers appear to have an edge in solving non-routine problems in physics, chemistry and mathematics. The objective of this study was to examine the spatial-problem solving relationship by observing engineering students while they solved word problems in mathematics, problems of some relevance to engineering education. The key findings from this work are highlighted in the first part of this chapter and reviewed against the literature to determine the extent to which they are compatible with previous research and have progressed the discussion on the role of spatial ability in the STEM curriculum.

Studies of this kind should not be conducted in isolation of engineering education but should be relevant to curriculum design, learning and teaching methods or in some other way improve how we educate engineers. Therefore, implications for curriculum design based on this work are suggested in the second part of the chapter. A mixed methods design was enacted by first testing a hypothesis using a quantitative approach and then using qualitative analysis to interpret approaches to problem solving. Others would have conducted this study in a different way and an evaluation of the research design along with some suggestions for future work in the area are also provided in the second part of the chapter before finishing with some concluding remarks.

8.1 Summary and analysis of findings

This study revealed several findings. First, spatial ability cannot be assumed to improve over time for those taking third level engineering courses of study. Second, spatial ability can play an important role in achieving particular learning outcomes but this can be hidden by correlations with course grades. Third, as illustrated by the correlation with DIRECT, spatial ability plays an important role in completing tasks related to interpreting circuit diagrams and drawings. Fourth, the relationship between spatial ability and problem solving is due to the representation and not the solution step for engineering students working on math story problems. Fifth, when a problem can be solved without representation the correlation with spatial ability disappears. Sixth, problems whose representation is driven by linguistic translation correlate with spatial ability but less so than for problems where the representation challenge is schematic. Seventh, competence in correctly representing problems is always accompanied by strong spatial ability and lack of such competence is nearly always accompanied by weak spatial ability. Each of these findings is discussed below with respect to the literature.

Spatial ability data collected from students in all four years of electrical engineering at DIT showed there to be no change in spatial ability with years spent on the course/programme. While a longitudinal study tracking the same sample over time or observations of a different curriculum might produce a contradictory result, this finding does suggest that studying engineering does not necessarily lead to improvement in spatial ability. Students can pass through an engineering programme and leave with the same level of spatial ability as they had when they joined, at least for those aspects of spatial ability that are measured by the PSVT:R and MCT. One would hope the same does not apply to the two other primary abilities, verbal and mathematical! Given that spatial ability can predict the highest level of engineering education achieved (Wai et al., 2009), one's engineering education fate appears to be significantly determined before the first day in engineering school. Intelligence, both fluid and crystallised, as described by the Cattell-Horn theory (Sternberg, 2004a) should be increased by studying engineering whereas this finding suggests that an important aspect of fluid intelligence can remain unchanged. Those who wish to see engineering students become more prepared for post graduate studies should reflect on the extent to which the curriculum is developing all abilities through all years of study.

Correlations between spatial ability and some learning outcomes can be hidden behind small or insignificant correlations with course grades. Clearly, the ability to interpret circuit diagrams shares common ground with spatial ability, as revealed by the large correlation between the two measures, yet grades in the circuits course had a much lower correlation with spatial ability. Spatial ability appears to be important and unimportant at the same time. Surely electrical engineering is more enjoyable for those who are more adept at interpreting the well-structured images contained in circuit diagrams. Perhaps higher levels of spatial ability accompany higher levels of self-efficacy in such areas. Both this and the previous finding suggest that engineering curricula that reward knowledge of facts, rote learning and memorisation are unlikely to reveal correlations with spatial ability. One could argue that a curriculum that does not require spatial thinking is one where problem solving, conceptual understanding and fluid thinking are not valued in assessment practice. It may be important in some areas of such curricula but will be hidden by overall correlations and will, therefore, remain a cognitive factor of little importance unworthy of development.

Of the four groups of questions on DIRECT, the correlation with spatial ability was due to Group A questions which required the interpretation of circuit diagrams. Visualizing and transforming well-structured images such as these diagrams is the very definition of spatial ability as described by (Linn & Petersen, 1985, p. 1482): "*Spatial ability generally refers to skill in representing, transforming, generating, and recalling symbolic non-linguistic information*". A feature of engineering practise is the extent to which sketches, drawings and symbols are used to communicate ideas and designs and this finding could be interpreted to mean spatial ability accompanies performance in these tasks also. DIRECT Group A questions are very different to the math problems that also correlated significantly with spatial ability. This suggests that spatial ability is manifest in two ways – visualizing and transforming text-free, well-structured image tasks on the one hand and image-free, word problem representation tasks on the other. How can differences in performance on a test such as the PSVT:R be connected with variation in performance on both types of task?

Lohman (1993) listed four possible explanations for difference in performance on tests of spatial ability – differences in mental rotation speed, working memory capacity, mental image generation ability and ability to process images holistically rather than analytically. The latter two explanations seem appropriate in the case of the DIRECT Group A questions. DIRECT is not regarded as a speeded test as most participants have more than enough time to complete all questions in the time allowed. To answer Question 27 on DIRECT (Figure 4-3), the challenge is to visualise each sketch, mentally trace the route of the wire and check that it travels through the battery and through the bulb filament. Assuming one has suitable semantic knowledge of the battery and the bulb, the challenge is one of image processing. Working memory also provides a plausible explanation for variation in performance and is discussed below after discussing the findings from the math problems and questions tests.

Findings from the quantitative analysis of the math problems and questions tests clearly show that spatial ability is most relevant to the problem representation step and this is supported by the qualitative analysis of the participants' solutions to the problems. A strong correlation was measured for the spatial-problem solving relationship while nothing of significance was found for the spatial-problem solution relationship. Spatial ability matters when problem translation and schema identification are required. These findings support those of Hegarty & Kozhevnikov (1999) who analysed sketches produced by 6th grade children while they solved math story problems and divided them into schematic and pictorial sketches. In both this and the Hegarty & Kozhevnikov (1999) study, participants who demonstrated appropriate schematic representations of the problems were more likely to be strong visualizers. A more open approach to categorising problem representation was taken in this study in that participants were free to demonstrate a representation in whatever way they chose. Some provided sketches, others did not but none were forced to do so. The analysis in this study revealed that participants do not have to commit their schemata to paper in the form of sketches in order to demonstrate what they are and they can still form appropriate schemata without having to sketch which may be a more natural approach for some. Boonen, van Wesel, Jolles, & van der Schoot (2014) conducted a similar study, also with 6th grade students, and they too found a significant relationship between spatial ability and ability to produce accurate visual-schematic representations ($r(126) = .31, p < .01$). The nature of the spatial-problem representation relationship defined by Boonen et al. (2014) is very similar to that described in this study between spatial ability and ability to appropriately represent the five problems (excluding the Pencils & Jars problem) for which the correlation was measured to be $r(113) = .585, p < .01$. A comparison of these two correlations shows the spatial-problem representation relationship is established by 6th grade and, for whatever reason, is larger among freshman engineering students.

Reasons for the difference in correlation reported by Boonen et al. (2014) and in this study are either due to research design – methods, samples, instruments – and/or difference in maturity of the participants. First, the samples differ in two ways – age and cognitive ability heterogeneity. The 6th grade students were aged 11 to 12 years old, the sample contained an equal number of boys and girls, and was purposely selected to give heterogeneity in mathematical ability. Likewise, the nature of the sample of school children suggests heterogeneity with regard to spatial ability also. The freshman engineering students from DIT and OSU in this study were aged 18 to 20 and, given the high standard of math required for acceptance to their courses, could be considered to come from a less heterogeneous group in terms of math ability. While the findings of Wai et al. (2009) suggest that the freshman engineering sample is also more restricted with regard to spatial ability, that does not hold in this case as the OSU portion of the sample was purposely selected to include a majority of weak visualizers. Hence, PSVT:R data from the full sample, DIT and OSU combined, was found to have a large standard deviation (M = 19.70, S.D. = 5.613).

A second reason for the difference may be related to instruments – both spatial and mathematical. Different tests were used to assess spatial ability: the PSVT:R was used in this study while Boonen et al. (2014) used two spatial tests in their study, the Paper Folding Task and the Picture Rotation Task. However, the same spatial factor, spatial visualization, was measured in both cases. Since freshman engineering and 6th grade students vary in semantic and subject prior knowledge, different types of math word problems were used in each study. Boonen et al. (2014) used the Mathematical Processing Instrument (MPI) compiled by Hegarty & Kozhevnikov (1999) which contains 14 word problems that mostly require basic arithmetic for problem solution. One requires a rectangle area schema but knowledge of volume, properties of a circle or algebra is not required. The problems used in this study are slightly more difficult but presented in a similar word problem style. Question 9 on the MPI is very similar to the Cans problem used in this study. Both the spatial and math tests are comparable for each study but the samples differ greatly in age, semantic knowledge, math ability, spatial ability and heterogeneity in math ability.

A third explanation may lie in the different maturity levels of the two samples. Spatial ability, like other human abilities, develops with age. For example, some tests, such as the MCT, are very unreliable when used with middle school children (Hungwe et al., 2014) but are reliable when used with engineering students (Sorby & Baartmans, 2000). This suggests a noticeable difference in spatial ability between the two age groups. Children in 6th grade are cognitively different to freshman engineering students and findings from studies of cognitive abilities with one group cannot necessarily be applied to another. An opportunity for spatial development exists between age 12 and 18 that will be grasped by some more than others. For example, some will have hobbies or engage in school activities that develop spatial ability while others won't. Variation in spatial ability among the population will therefore grow during these years. Relative to the 6th grade sample, the freshman engineering sample in this study may contain a larger variation in spatial ability which can explain some portion of the increase in correlation.

As illustrated by solutions to the Pencils & Jars problem, when a problem can be solved without representation, success in solving the problem becomes independent of spatial ability. This finding simply reaffirms the notion that problem solving consists of representation plus solution and the relationship with spatial ability is accounted for by the representation phase because in cases such as this where representation is not necessary success is unrelated to spatial ability. If administered differently by requiring participants to use an algebraic approach, data collected from Pencils & Jars problem would have provided some insight into the relationship between spatial ability and success in translating a word statement to an algebraic equation. The few who were successful in making this translation also had high scores on the PSVT:R but there were not enough cases to pass a test of significance. Clement (1982) found that only 27 % of the sample in his study were successful in translating the word statement from the Cheesecake & Strudel problem to an algebraic equation. Findings from this study suggest this group contained more strong visualizers that the group that was unsuccessful. Resolving this issue is left to future work.

The Pencils & Jars problem challenged the definition of problem representation used in this study because it showed there are times when a representation that can be labelled as mathematically inappropriate is successful in guiding the problem solution. This problem revealed there can be multiple valid representations for the same problem and the relationship between spatial ability and representation depends on the nature of the representation. In other words, some representations reveal greater overlap with spatial ability than others. This, in turn, challenges the finding that spatial ability is significantly related to problem representation for if the latter can vary in form then the relationship may not always hold. Hence, a more inclusive definition of the type of problem representation that is related to spatial ability is needed. Informed by the information processing model of intelligence, it is speculated that spatial ability is related to problem representation only when representation places a high demand on working memory capacity and on the visuospatial aspect of working memory in particular. Sometimes, representation that leads to solution can be achieved with little demand on working memory as in the case of guess and check, for example. In contrast, translating word statements to algebraic equations places a large demand on working memory and, therefore, this form of representation is related to spatial ability.

Without data from the Pencils & Jars problem the finding of this study would simply be a strong relationship between spatial ability and appropriate representation of word problems in mathematics. To allow for data collected from Pencils & Jars, a more nuanced definition of the spatial-representation relationship is required as this relationship only holds in certain cases as possibly explained by representation placing high demands on working memory capacity. Hence, the definition of spatial ability provided earlier is now adjusted to account for this as follows: spatial ability is the ability to mentally organise information from sources that vary from well-structured images to word descriptions of quantities, that are sufficiently novel or non-routine to place a high demand on working memory, and rearrange this information into a new format that is consistent with the original form and to do so promptly.

Following Mayer's framework, problem representation was considered in this work to draw on three different types of knowledge – linguistic, semantic and schematic. For two problems, success in the representation step hinged on linguistic ability and success in this regard was found to be significantly related to spatial ability with a medium to large effect size. Where success in representation hinged on schematic knowledge, the correlation was also significant but with a much larger effect size. Why does spatial ability appear to play a greater role when the representational challenge is schematic rather than linguistic? Perhaps linguistic representation depends more on verbal ability than schematic representation and perhaps schema are developed with visual imagery.

Lohman (1993) placed much emphasis on a working memory explanation for differences in spatial ability as he saw enough overlap between spatial ability, working memory and 'g' to speculate they may all be the same thing. His theory can be applied to the findings of this study as follows. Working memory is arguably stretched to its limits during problem translation and representation so those with greater working memory capacity make fewer mistakes because they can concurrently process more pieces of information. It is possible that spatial ability tests serve to rank the sample by the visuospatial sketchpad component of working memory capacity while tests of problem solving rank the sample by both components of working memory, hence the large correlation or shared variation between the two sets of scores. Problem solving requires both components of working memory – phonological loop and visuospatial sketchpad – whereas spatial tests mostly assess the visuospatial sketchpad aspect. DIRECT Group A questions mostly draw on this aspect of working memory, hence the correlation. Linguistic representation may be more determined by the phonological loop which is not assessed by the PSVT:R, hence the lower correlation. Schematic representation draws on both aspects of working memory and this explains its larger correlation. Tests of core competency assess long term memorised knowledge of math procedures which is possessed by the majority in the sample, hence variation in working memory capacity, as measured by the PSVT:R, is not reflected in the core competency scores.

Finally, the scatter plots of spatial ability versus problem score and versus problem representation score indicate that one cannot be good at problem representation unless one possesses strong spatial ability; one who is good at problem representation is very likely to perform well on test of spatial ability such as the PSVT:R, i.e. the former is an accurate prediction of the latter. However, there are also those who have strong spatial ability but did not perform well at problem representation which indicates that other factors are involved in addition to spatial ability when representing problems. These may include as prior knowledge, self-efficacy, motivation and personal epistemology. Spatial ability appears to be an essential but not the only ingredient. In addition, if visuospatial working memory is manifest as spatial ability then one can rephrase the statement to say that problem representation is facilitated by visuospatial working memory.

8.2 Spatial ability vs. other abilities

Spatial ability is the only cognitive factor that was measured and compared with the problem solving process in this study. Although mathematical ability was accounted for through the core competency questions it was only in this context that it was measured rather than as a reliable and valid measurement of this ability. The role played by mathematical and verbal abilities in solving these problems remained unknown and, to play the devil's advocate, these abilities may be equally or more important to the cognitive activities involved in solving these problems as spatial ability was shown to be. In her meta-analysis, Friedman (1992) found verbal ability to have a significant relationship with mathematics and, given these problems are mathematical, it is plausible that verbal ability may have been very relevant. Furthermore, as argued by Lohman (1993), tests of spatial ability have been shown to be very good measures of general intelligence, i.e. that aspect of an individual's cognitive makeup that determines performance on all tests of thinking and it is therefore possible that the relationships discussed here between spatial ability and problem solving/problem representation are really relationships between general intelligence and problem solving and representation. To highlight the purely spatial aspect of these relationships would require data from other psychometric tests that measure mathematical and verbal ability and/or general intelligence. If one found that another test of intelligence was equally well related to problem solving and representation as spatial ability then one could conclude that spatial tests are simply ranking the sample by general ability rather than a unique spatial ability factor of intelligence. On the other hand, if the spatial/problem solving correlation was found to be significantly stronger than the general intelligence/problem solving correlation one could then conclude that problem representation and solving truly does draw on spatial ability.

It was possible to retrospectively examine this issue to some extent as SAT and ACT data had been collected for the studies presented in Chapter 4 and were, therefore, available for some of the OSU participants. ACT math, english, reading and science reasoning (SCIRE) data were available for 31 of these students while SAT math data were available for 4 students and were converted to ACT math scores using a conversion table (Grove, n.d.). A correlation matrix was created to separately measure the relationship between each of these variables as shown in Table 8-1.

	ACT Math	**ACT English**	**ACT Read**	**ACT SCIRE**	**Problem score**	**Problem representation**
PSVT:R	.249 (35)	.159 (31)	-.275 (31)	-.180 (35)	.577** (115)	.585** (115)
ACT/SAT Math		-.020 (31)	-.098 (31)	.047 (31)	.441** (35)	.289 (35)
ACT English			.345 (31)	-.006 (31)	-.065 (31)	-.120 (31)
ACT Read				.143 (31)	-.137 (31)	-.111 (31)
ACT SCIRE					-.087 (31)	-.114 (31)
Problem score						.715** (120)

** significant at $p < .01$

Table 8-1. Correlation matrix for all students for whom data were available. The number of cases used is shown in brackets after each correlation value. Problem score and representation are for 5 problems, i.e. Pencils & Jars is excluded.

Mathematical ability, as measured by the SAT math and ACT math tests was found to be significantly related to performance in problem solving but not to problem representation and, relative to spatial ability, with smaller effect sizes. Verbal ability, as measured by the ACT English and reading tests was not found to be related to problem representation or solving. Likewise, the ACT science reasoning test was not significantly related to either aspect of problem solving. It appears that in solving the simple math word problems used in this study, both mathematical and spatial abilities are relevant but not verbal ability and that of these, spatial ability has a slightly larger effect size. In terms of representing the problems, however, only spatial ability was found to be relevant marking it out as separate and distinct from the other two abilities in this thought process.

A significant relationship between math ability and problem solving is to be expected as the problems are mathematical in nature. As tests of mathematical ability, the SAT and ACT have been found to have high reliability and validity and to be very good predictors of success in higher education in the US (Camara & Echternacht, 2000; Powers et al., 2016). One could regard them as measures of individual abilities and as metrics of general intelligence. It is interesting to find relationship with problem representation being different for spatial ability and SAT/ACT math. Excluding the Pencils & Jars problem, i.e. for the other five problems, success in problem solving depended on having an appropriate representation of the problem and on correctly carrying out the solution steps. Problem solving scores reflect the successful completion of both of these steps whereas problem representation scores reflect the successful completion of the first step only. Since SAT/ACT math and spatial ability are significantly related to the combination of representation and solution one could conclude spatial ability is acting as a measure of general intelligence in this case. For problem representation, however, SAT/ACT math is not significantly related to it whereas spatial ability is and with an even stronger effect size compared with problem solving. In this case, therefore, spatial ability appears to measure an aspect of thinking that is not measured by math ability tests.

This conclusion is supported by the lack of significance in the relationship between problem solving/representation and the other ACT measures of reading, English and science reasoning, i.e. representing the math story problems used in this study required an aspect of thinking that is not measured by ACT math and verbal tests but is required by the PSVT:R, a test of spatial ability. In this case, problem solving appears to draw on math and spatial abilities but not on verbal ability while problem representation appears to draw on an aspect of thinking that is also measured by a test of spatial ability and appears to be unique to this ability and separate to general intelligence.

Friedman's (1992) finding that verbal-mathematical correlations were either equal to or higher than spatial-mathematical correlations is not supported in this case. It is possible these problems require a level of verbal ability that is well exceeded by this sample of first year college students. If the same problems were administered to younger samples, as many in Friedman's meta-analysis were, a more significant relationship with verbal ability might emerge as some may fail to understand the problems because of less poorly developed verbal skills. In that case, problem solving might test verbal, mathematical and spatial abilities.

It is regrettable that verbal and mathematical ability data were not collected from all the participants as the above discussion is based on a limited set of data and the findings may be different if a full set of SAT/ACT data were available. Directly collecting these data would have required the participants to contribute significantly more of their time to the study, something which had not been allowed for in the design. In addition, the SAT and ACT tests are not administered in Ireland where state examinations are used to assess mathematics and English in a style that is sufficiently different to make conversion from one to the other unreliable. It is therefore left to future work to further evaluate the relationship between problem representation and all primary abilities.

8.3 Evaluation, relevance and future directions

Engineers are reputed to be good problem solvers; they provide the bridge (literally and figuratively) between current and desired states. Engineers Ireland (EI), the accrediting body for engineering education in Ireland, includes "*the ability to identify, formulate, analyse and solve complex engineering problems*" as one of its key engineering programme outcomes

(Owens, 2014, p. 16). ABET and Engineers Australia, the equivalent bodies in the US and Australia, contain similar statements in their criteria for engineering education to achieve. Problem solving is an essential feature of engineering and, based on findings from this study, spatial ability is an essential component of success in problem solving. Yet, it is possible to find no differences in spatial ability between samples of students in all years of an engineering programme. This raises a question for the engineering education community: if problem solving is important in engineering education and spatial ability can play an essential role in the problem solving process why does spatial ability development not appear on the curriculum?

Problem solving has been shown in this study to consist of two cognitively distinct phases – representation and solution – with students having the most difficulty in the representation phase where spatial ability was most relevant. Implications for learning, teaching and assessment (LTA) activities include separation at the learning outcome level of problem representation and solution, direct intervention to improve problem solving strategies and indirect intervention to improve cognitive development in spatial ability.

Given students struggled to represent these problems but were competent in the mathematical procedures required, one could consider redesigning learning outcomes and LTA to reflect this difference. Procedural knowledge could be developed through self-directed activities, facilitated through online learning, for example, in which the student at her own time works through a series of activities that demonstrate procedural knowledge and skills. The many rules and laws that are covered in different STEM courses could be addressed in this way. Online assessment with formative feedback could be designed to ensure these learning outcomes are met and procedural knowledge continues to be acquired. Class time could then be devoted to problem solving activities in which the formal learning outcomes addressed relate to issues of problem representation. Such an approach would be very compatible with a problem or project-based learning approach in which students are required to engage with solving problems that are challenging, complex and relevant. Dividing problem solving into these two steps, or even dwelling on the distinction for a while, might open up some opportunities to address the part that really challenges students, problem representation.

Direct intervention to develop problem solving skills in a general way that has relevance not just to one course or module but to problem solving in any context should also be considered based on these findings. Many models, tips, techniques and advice on problem solving methods are easily obtained and applied and instructors should, and indeed, many do, consider giving a small amount of time to such activities in their timetables. Critical thinking techniques such as those developed by De Bono (2010), for example, would greatly support such strategies and provide engineering students with methods for approaching problem solving they could apply in many contexts including professional life. Both of these activities could be incorporated into existing courses without much difficulty.

Indirect intervention in the form of spatial skills training should also be considered as it offers the possibility that developing this aspect of cognition could lead to benefits in many areas including problem representation ability. Correlation does not equal causation, however, and the findings in this study do not indicate that attending such a course would lead to improvements in representation. Nor do the findings rule out this possibility. Logic indicates that enhancing spatial ability, a primary cognitive factor of intelligence, will lead to improved performance on spatial tasks, one of these being problem representation. Sorby and colleagues (Sorby & Baartmans, 2000; Veurink & Sorby, 2011) have measured improvements in course grades and retention rates among weak visualizers who attended a one semester spatial skills training course. In addition, if spatial ability is explained by working memory then courses to develop spatial ability could be evaluated based on the extent to which they address improvements in working memory.

How would these arguments be received by management in engineering schools? Would the department chair or head of school readily agree to remove an existing course or module from the curriculum and replace it with spatial skills training? A sceptic might not be convinced based on the evidence from this study and would likely need proof of causation from quantitative data before committing resources. Others have been willing to try based on the evidence provided by Sorby and colleagues (Sorby & Baartmans, 2000; Veurink & Sorby, 2011) and findings from this study could assist in convincing those amenable to reform. Changes in approach to problem solving following spatial skills training could be evaluated in future work.

A mixed methods design was employed in this study which allowed the spatial-problem representation relationship to be identified with a large degree of certainty and which was then analysed through an interpretive approach to yield descriptions that are claimed to have a large element of truth as they can be traced directly back to what the participants produced. The project made several achievements that have allowed the discussion in the literature about the spatial-problem solving relationship to progress to some extent. A reasonable argument for generalizing these findings can be made. The sample consisted of engineering students from two locations in different countries thereby broadening the contextual base of the study. The problems required little discipline specific knowledge and when participants were excluded for not having this knowledge the relationship between the two variables remained significant. The problems were more relevant to engineering than others that have been used in studies of this nature such as river crossing problems (missionaries and cannibals) and this should make the work more relevant to the engineering education community, an important audience for this work. Rather than apply the philosophy of the BFG, a conscious effort was made to align epistemology, theoretical perspective, methods and research questions.

There are several potential drawbacks to this work. To some participants, the problems may have appeared to be exercises, i.e. they possessed prior knowledge of how to solve some or all of the problems. There are likely many other reasons why the solutions of all 115 participants to all 6 problems may not have been true examples of problem solving. Although the think aloud data could be analysed for signs of ease in problem solving such as very short pauses between writing that might indicate the problem is an exercise, such recordings exists for only one quarter of the data set. Reliability of the two math tests was found to be poor which indicates there are other factors at play and those with a positivist perspective are likely to criticise this aspect. The PMI used by Hegarty & Kozhevnikov (1999) was found to be reliable for their sample. The tests could be redesigned to improve internal consistency by reducing the range of problem type, for example. Removing the Pencils & Jars problem would probably improve internal consistency but would also remove a key finding from the study. Rather than view the problems as a test, the individual problem analyses could be emphasised in reporting the work. Positivists are also likely to take issue with the lack of an inter-rater reliability score for the interpretations of problem solutions. However, this was an individual research project and no claims are made that a conclusive answer has been provided - that is left to the reader to judge. Composite scores from several spatial tests are favoured by some researchers (e.g., Lohman, 1993) but, for practical reasons, primarily participant time, only one test was administered in this case along with the two math tests. However, as shown in Chapter 5, the spatial-problem representation correlation was repeated with another sample based on MCT data.

Interpretivists may lament the emphasis on prior assumptions that came with spatial ability as a factor of intelligence and criticise the omission of affective issues such as self-efficacy, motivation and personal epistemology in the research design. Significantly more time and resources would have been required to incorporate these aspects than were available for the project but they merit attention in future work. Paper and pencil tests for self-efficacy, motivation and personal epistemology exist and could be administered in a follow up study. Phenomenologists will note the absence of epoché in the research design, a consequence of substantial prior assumptions. 'I wouldn't start from here', one can hear them say.

Future work should examine problem solving in other subject areas. For example, a similar study could be conducted using electric circuit problems given the relationship that was discovered between spatial ability and understanding of physical aspects of simple DC circuits. It would be interesting to see if the same problem representation/solution divide would emerge. Electric circuits word problems in which linguistic and schematic representations are obscured would be required along with a test of basic competencies in the subject to rule out procedural knowledge as a third variable. Likewise for problems taken from statics and other core areas of engineering curricula. If similar finding emerged, weight would be added to the argument for the relevance of spatial ability to engineering education.

The relationship between spatial ability and problem solving exposed in this study is itself a source of insight into each variable in this relationship. While the definition of spatial ability provided at the beginning of the study resonates well with the way spatial ability is measured by the PSVT:R it fails to account for its relationship with problem representation. Essential elements of this definition - the well-structured image, visualizing transformations of and between two and three dimensional images - are missing from the word problems. The findings of this study show there are cognitive processes that are common to mentally rotating the PSVT:R images and mentally representing the word problem statements. Either spatial ability is something more than mental visualization and transformation of well-structured images or there is another ability that is related to both processes. If the latter then what is this other ability? Study after study over the past hundred years has pointed to only three primary abilities – spatial, mathematical and verbal. It seems logical to reflect on the definition of spatial ability rather than search elsewhere. This definition may need to be sufficiently broad or inclusive to cover tasks that range from tasks associated with problem representation to mental visualization and transformation of well-structured images.

Working memory, a component of the central executive in information processing models of intelligence that has been shown to have a spatial component, offers a way of explaining the relationship between performance on the PSVT:R and problem representation. Assuming both tasks are novel to the participants, long term prior knowledge is not available to assist so the participant is heavily dependent on working memory capabilities. This line of argument suggests a key finding from this work is that an essential feature of a task for it to relate to spatial ability is that it is novel or non-routine to the participant and when this is not the case spatial ability is less relevant. Perhaps the relevance of spatial ability in problem representation is linked to imagery: when processing novel tasks there is no pre-existing, readily available imagery to assist in the creation of a representation; the novelty requires the creation of imagistic representation and this process is related to spatial ability. Hence, findings from this study suggest a revised definition of spatial ability: spatial ability is the ability to mentally organise information from sources that vary from well-structured images to word descriptions of quantities, that are novel or non-routine, create visual representations from this information that allow it to be organised into a new format that is consistent with the original form and to do so promptly.

8.4 Contribution to knowledge

1. There is a very strong and significant relationship between spatial ability and performance in solving simple word problems in mathematics. This relationship appears to be entirely based on the problem representation step and is independent of the solution step.

2. As spatial ability was related to problem representation and not problems solution, for the sample of 1st year engineering students used in this study, solving word problems in

mathematics consists of two phases – representation and solution – that have been shown in this study to be cognitively distinct. Representation can be heavily dependent on spatial ability, an aspect of cognition that is irrelevant to the solution phase.

3. It appears that if the cognitive demands of problem representation can be reduced by the use of guess and check, for example, then its relationship with spatial ability becomes insignificant.

4. Those with high levels of spatial ability are much more likely to appropriately represent and solve word problems in math. STEM educators who seek to develop the ability to represent problems similar to those used in this study cannot ignore spatial ability as a key component of the student's intellectual makeup. While correlation does not equal causation, these findings challenge educators to think more broadly than the traditional view that only mathematical and verbal abilities are formally developed in the STEM curriculum.

5. As shown in Tables 7-10 and 7-11, schematic representation ability appears to be more strongly related to spatial ability than linguistic representation ability. This may be explained by the distinction in working memory between phonological and visual tasks: both the PSVT:R and schema development are more dependent on the latter than the former while linguistic representation also draws on the former, the phonological loop.

6. The relationship between spatial ability and problem representation shown to be established among 6th grade students at age 12 (Hegarty & Kozhevnikov, 1999), is also present among a sample of first year engineering students that has completed a further 6 years of post-primary education.

7. Spatial ability is not limited to visualization and transformation of well-structured images. In the context of STEM education at least, it is the ability to mentally organise

information from sources that vary from well-structured images to word descriptions of quantities, that are sufficiently novel or non-routine to place a high demand on working memory, and rearrange this information into a new format that is consistent with the original form and to do so promptly.

8. As measured by the PSVT:R, spatial ability is highly relevant to the analysis of simple DC electric circuits, a skill/ability that electrical engineering students or any students who engage in circuit design, analysis and construction, should possess. This finding has relevance for those who promote STEM education and careers through electric circuit construction activities.

9. Spatial ability cannot be assumed to develop during four years of full time study of electrical engineering.

10. That spatial ability is not necessarily developed through studying electrical engineering but is highly relevant to a core aspect of electrical engineering is a contradiction that can be explained by the assessment process. Assessment is sufficiently varied that one can still succeed overall while underperforming in certain areas.

Concluding remarks

There is a growing awareness among the engineering education community that curriculum design should be informed by research so that design decisions are based on evidence from rigorous research studies. It has been shown in this study that spatial ability is an aspect of cognition that is highly relevant to assessments that require reasoning about concepts, novel scenarios and problems. If engineers are to develop these skill and become good problem solvers then the findings of this study can be used to justify the inclusion of spatial ability assessment and development in the engineering curriculum. Another highlight of this work is that cognitive performance was found to be markedly different in each phase – representation and solution – of the problem solving process. Therefore, the learning, teaching and assessment activities for problem solving should be reviewed with this in mind as each phase arguably requires a different strategy. Engineering educators should reflect on this finding and consider ways to help those students who are particularly challenged by the representation phase to become better problem solvers.

Bibliography

Andrews McMeel Syndication - Home. (2017). Retrieved April 5, 2017, from http://syndication.andrewsmcmeel.com/licensing_permissions/educational_use

Anon. (n.d.). http://www.projecttalent.org/.

Ashworth, P., & Greasley, K. (2009). The phenomenology of “approach to studying”: the idiographic turn. *Studies in Higher Education*, *34*(5), 561–576. https://doi.org/10.1080/03075070802597184

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Academic Press. https://doi.org/10.1016/S0079-7421(08)60452-1

Bodner, G. M. (2015). Research on Problem Solving in Chemistry. *Chemistry Education: Best Practices, Opportunities and Trends*, 181–202.

Bodner, G. M., & Guay, R. B. (1997). The Purdue visualization of rotations test. *The Chemical Educator*, *2*(4), 1–17.

Boonen, A. J. ., van Wesel, F., Jolles, J., & van der Schoot, M. (2014). The role of visual representation type, spatial ability, and reading comprehension in word problem solving: An item-level analysis in elementary school children. *International Journal of Educational Research*, *68*, 15–26.

Borrego, M., Douglas, E. P., & Amelink, C. T. (2009). Quantitative, Qualitative, and Mixed Research Methods in Engineering Education. *Journal of Engineering Education*, *98*(1), 53–66.

Branoff, T. J. (1998). *The effects of adding coordinate axes to a mental rotations task in measuring spatial visualization ability: An information-processing approach relating to teaching methods of undergraduate technical graphics education*. North Carolina State University. Retrieved from http://elibrary.ru/item.asp?id=5494485

Camara, W. J., & Echternacht, G. (2000). The SAT I and High School Grades: Utility in Predicting Success in College. Retrieved from https://research.collegeboard.org/publications/content/2012/05/sat-i-and-high-school-grades-utility-predicting-success-college

Carroll, J. B. (1996). A three-stratum theory of intelligence: Spearman's contribution. In I. Dennis & P. Tapsfield (Eds.), *Human abilities: Their nature and measurement* (pp. 1–17). Retrieved from https://books.google.com/books?hl=en&lr=&id=JQX-PlF6EgYC&oi=fnd&pg=PA1&dq=Three-Stratum+Theory+of++intelligence&ots=_pNLaysHXX&sig=fm0uwps6DsjFEyTcAJsDfFPNoPQ

Carter, C. S., LaRussa, M. A., & Bodner, G. M. (1987). A study of two measures of spatial ability as predictors of success in different levels of general chemistry. *Journal of Research in Science Teaching*, *24*, 645–657.

Casey, M. B., Nuttall, R. L., & Pezaris, E. (2001). Spatial-mechanical reasoning skills versus mathematics self-confidence as mediators of gender differences on mathematics subtests using cross-national gender-based items. *Journal for Research in Mathematics Education*, 28–57.

Casey, M. B., Nuttall, R., Pezaris, E., & Benbow, C. P. (1995). The influence of spatial ability on gender differences in mathematics college entrance test scores across diverse samples. *Developmental Psychology*, *31*(4), 697–705. https://doi.org/10.1037/0012-1649.31.4.697

CEEB. (1939). *CEEB Special Aptitude Test in Spatial Relations*. New York: College Entrance Examination Board.

Choudhary, N., & Jesiek, B. K. (2016). State of Mixed Methods research in Engineering Education: In-depth examination of JEE articles, 2010 #x2013;2015. In *2016 IEEE*

Frontiers in Education Conference (FIE) (pp. 1–9). https://doi.org/10.1109/FIE.2016.7757554

Clement, J. J. (1982). Algebra Word Problem Solutions: Thought Processes Underlying a Common Misconception. *Journal for Research in Mathematics Education*, *13*(1), 16–30.

Clement, J. J., Soloway, E., & Lochhead, J. (1982). Does computer programming enhance problem solving ability?

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, N.J.: L. Erlbaum Associates.

Coquin-Viennot, D., & Moreau, S. (2003). Highlighting the role of the episodic situation model in the solving of arithmetical problems. *European Journal of Psychology of Education*, *18*(3), 267–279.

Creswell, J. W. (2003). *Research design: qualitative, quantitative, and mixed method approaches*. Thousand Oaks, Calif.: Sage Publications.

Crotty, M. (1998). *The foundations of social research : meaning and perspective in the research process*. London: SAGE.

Dahl, R. (1984). *The BFG*. Harmondsworth: Puffin in association with Cape.

Danielak, B. A., Gupta, A., & Elby, A. (2014). Marginalized Identities of Sense-Makers: Reframing Engineering Student Retention. *Journal of Engineering Education*, *103*(1), 8–44. https://doi.org/10.1002/jee.20035

Davis, B. (2015). *Spatial reasoning in the early years: Principles, assertions, and speculations*. Routledge. Retrieved from https://books.google.com/books?hl=en&lr=&id=5H9KCAAAQBAJ&oi=fnd&pg=PP1&dq=spatial+reasoning+in+the+early+years&ots=kSloZXsyxn&sig=aWlYycvfhUAnWCusiut3v-KcvJs

De Bono, E. (2010). *Teach yourself to think* (New ed). London: Viking.

Dennis, I., & Tapsfield, P. (Eds.). (1996). *Human abilities: their nature and measurement*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Doyle, R. A., Voyer, D., & Cherney, I. D. (2012). The relation between childhood spatial activities and spatial abilities in adulthood. *Journal of Applied Developmental Psychology*, *33*(2), 112–120. https://doi.org/10.1016/j.appdev.2012.01.002

Duffy, G., Farrell, S., Harding, R., Nevin, E., Mac Raighne, A., Howard, R., & Bowe, B. (2015). The effects of spatial skills and spatial skills training on academic performance in STEM education. Presented at the Research in Engineering Education Symposium.

Duffy, G., Sorby, S. A., Mack, A., & Bowe, B. (2017). Performance by Gender on University Placement Tests in Mathematics and Spatial Skills. Presented at the ASEE Annual Conference 2017, Columbus Ohio: American Society for Engineering Education. Retrieved from https://peer.asee.org/28737.pdf

Duncker, K. (1945). *On problem-solving*. Washington, D.C.: American Psychological Association.

Edens, K., & Potter, E. (2008). How Students "Unpack" the Structure of a Word Problem: Graphic Representations and Problems Solving. *School Science & Mathematics*, *108*(5), 184–196.

Einstein, A. (1920). *Relativity: The special and general theory*. Penguin. Retrieved from http://www.gutenberg.lib.md.us/3/6/1/1/36114/36114-pdf.pdf

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). Manual for kit of factor-referenced cognitive tests. *Princeton, NJ: Educational Testing Service*.

Eliot, J., & Smith, I. M. (1983). *An international directory of spatial tests*. Cengage Learning Emea.

Engelhardt, P. V., & Beichner, R. J. (2004). Students' understanding of direct current resistive electrical circuits. *American Journal of Physics*, *72*(1), 98–115.

Eysenck, M. W. (2001). *Principles of cognitive psychology*. Hove, UK; Philadelphia: Psychology Press.

Farrell, S., Behan, A., Duffy, G., Harding, R., Howard, R., Mac Raighne, A., … Bowe, B. (2015). A Profile of the Spatial Visualisation Abilities of First Year Engineering and Science Students. Retrieved from http://arrow.dit.ie/engschcivcon/90/?utm_source=arrow.dit.ie%2Fengschcivcon%2F90&utm_medium=PDF&utm_campaign=PDFCoverPages

Field, A. (2013). *Discovering Statistics using IBM SPSS Statistics* (Fourth Edition). London: SAGE Publications Ltd.

Friedman, L. (1992). *A Meta-Analysis of Correlations of Spatial and Mathematical Tasks*. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED353270&site=ehost-live

Friedman, L. (1995). The Space Factor in Mathematics: Gender Differences. *Review of Educational Research*, *65*(1), 22–50.

Geiser, C., Lehmann, W., Corth, M., & Eid, M. (2008). Quantitative and qualitative change in children’s mental rotation performance. *Learning and Individual Differences*, *18*(4), 419–429.

Geiser, C., Lehmann, W., & Eid, M. (2006). Separating “Rotators” From “Nonrotators” in the Mental Rotations Test: A Multigroup Latent Class Analysis. *Multivariate Behavioral Research*, *41*(3), 261–293.

Grove, A. (n.d.). How Does an SAT Score Translate to an ACT Score? Retrieved October 6, 2017, from https://www.thoughtco.com/converting-act-scores-to-sat-scores-788710

Guay, R. (1976). *Purdue spatial visualization test*. Purdue University.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest*, *8*(1), 1–51.

Hammer, D., & Elby, A. (2002). On the Form of a Personal Epistemology. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal Epistemology: The Psychology of Beliefs About Knowledge and Knowing*. Mahwah, NJ: Erlbaum.

Hegarty, M., & Kozhevnikov, M. (1999). Types of visual–spatial representations and mathematical problem solving. *Journal of Educational Psychology*, *91*(4), 684–689. https://doi.org/10.1037/0022-0663.91.4.684

Hegarty, M., Mayer, R. E., & Green, C. E. (1992). Comprehension of arithmetic word problems: Evidence from students' eye fixations. *Journal of Educational Psychology*, *84*(1), 76–84. https://doi.org/10.1037/0022-0663.84.1.76

Hegarty, M., Mayer, R. E., & Monk, C. A. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *Journal of Educational Psychology*, *87*(1), 18–32. https://doi.org/10.1037/0022-0663.87.1.18

Hinsley, D. A., Hayes, J. R., & Simon, H. A. (1977). From words to equations: Meaning and representation in algebra word problems. *Cognitive Processes in Comprehension*, *329*.

Hungwe, K., Sorby, S. A., Molzon, R., Charlesworth, P., & Wang, M. (2014). Supporting the development of spatial visualization in middle grade and high school students. *J. Women Minorities Sci. Eng. Journal of Women and Minorities in Science and Engineering*, *20*(4), 379–393.

Irving, P. (2010). *A Phenomenographic Study of Introductory Physics Students: Approaches to their Learning and Perceptions of their Learning Environment in a Physics Problem-Based Learning Environment*. Dublin Institute of Technology, Unpublished doctoral dissertation. Retrieved from http://arrow.dit.ie/sciendoc/96/

Isaacson, W. (2007). *Einstein: his life and universe*. London: Simon & Schuster.

Jonassen, D. H. (2010). *Learning to solve problems: A handbook for designing problem-solving learning environments*. Routledge.

Kelly Jr, W. F. (2012). *Measurement of spatial ability in an introductory graphic communications course*. North Carolina State University. Retrieved from http://search.proquest.com/openview/4b52d198fb1c15fc051f2ecb7246f581/1?pq-origsite=gscholar&cbl=18750&diss=y

King, P. M., & Kitchener, K. S. (1994). *Developing reflective judgement : understanding and promoting intellectual growth and critical thinking in adolescents and adults*. San Francisco: Jossey-Bass Publishers.

Kozhevnikov, M., Motes, M. A., & Hegarty, M. (2007). Spatial Visualization in Physics Problem Solving. *Cognitive Science*, *31*(4), 549–579. https://doi.org/10.1080/15326900701399897

Kozhevnikov, M., & Thornton, R. (2006). Real-Time Data Display, Spatial Visualization Ability, and Learning Force and Motion Concepts. *Journal of Science Education & Technology*, *15*(1), 111–132. https://doi.org/10.1007/s10955-006-10361-0

Kuhn, T. S. (1996). *The structure of scientific revolutions* (3rd ed.). Chicago, IL: University of Chicago Press.

Kyllonen, P. C. (1996). Is working memory capacity Spearman's g. *Human Abilities: Their Nature and Measurement*, 49–75.

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, 1479–1498.

Lippa, R. A., Collaer, M. L., & Peters, M. (2010). Sex differences in mental rotation and line angle judgments are positively associated with gender equality and economic development across 53 nations. *Archives of Sexual Behavior*, *39*(4), 990–997.

Lohman, D. F. (1993). Spatial ability and g. In I. Dennis & P. Tapsfield (Eds.), *Human abilities: Their nature and measurement*. New Jersey: Lawrence Erlbaum Associates, Inc.

Love, P. G., & Guthrie, V. L. (1999). Understanding and Applying Cognitive Development Theory. *New Directions for Student Services*, (88), 1–107.

Mac Raighne, A., Behan, A., Duffy, G., Farrell, S., Harding, R., Howard, R., … Bowe, B. (2015). Examining the relationship between physics students' spatial skills and conceptual understanding of Newtonian mechanics. Presented at the Research in Engineering Education Symposium.

Maeda, Y., & Yoon, S. Y. (2011). Measuring Spatial Ability of First-Year Engineering Students With the Revised PSVT: R. Presented at the American Society for Engineering Education, American Society for Engineering Education.

Maeda, Y., Yoon, S. Y., Kim-Kang, G., & Imbrie, P. K. (2013). Psychometric properties of the revised PSVT:R for measuring first year engineering students' spatial ability. *International Journal of Engineering Education*, *29*(3), 763–776.

Mayer, R. E. (1983). *Thinking, problem solving, cognition*. New York: W.H. Freeman.

Mayer, R. E. (1992). *Thinking, problem solving, cognition*. New York: W.H. Freeman.

McGee MG. (1979). Human spatial abilities: psychometric studies and environmental, genetic, hormonal, and neurological influences. *Psychological Bulletin*, *86*(5), 889–918.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review*, *63*, 81–97.

Mustata, A. (n.d.). *Maths Circles Ireland 1st Year Handbook*. Retrieved from http://mathscircles.ie/wp-content/uploads/2012/10/Handbook.pdf

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, N.J.: Prentice-Hall.

O'Cathain, A., Murphy, E., & Nicholl, J. (2008). The quality of mixed methods studies in health services research. *Journal Of Health Services Research & Policy*, *13*(2), 92–98. https://doi.org/10.1258/jhsrp.2007.007074

Owens, D. (2014, April). Accreditation Criteria for Professional Titles. Engineers Ireland.

Perry, W. G. (1999). *Forms of intellectual and ethical development in the college years : a scheme*. San Francisco: Jossey-Bass Publishers.

Peters, M., Chisholm, P., & Laeng, B. (1995). Spatial ability, student gender, and academic performance. *Journal of Engineering Education*, *84*(1), 69–73.

Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse Mental Rotations Test: Different versions and factors that affect performance. *Brain and Cognition*, *28*(1), 39–58. https://doi.org/10.1006/brcg.1995.1032

Potok, C. (1995). *The Chosen*. Holt, Rinehart and Winston.

Powers, S., Li, D., Suh, H., & Harris, D. J. (2016, October). ACT Reporting Category Interpretation Guide Version 1.0. ACT.

Pribyl, J. R., & Bodner, G. M. (1987). Spatial ability and its role in organic chemistry: A study of four organic courses. *Journal of Research in Science Teaching*, *24*(3), 229–240. https://doi.org/10.1002/tea.3660240304

Robbins, T. W., Anderson, E. J., Barker, D. R., Bradley, A. C., Fearnyhough, C., Henson, R., … Baddeley, A. D. (1996). Working memory in chess. *Memory & Cognition Memory & Cognition*, *24*(1), 83–93.

Schommer, M. (1990). Effects of Beliefs About the Nature of Knowledge on Comprehension. *Journal of Educational Psychology*, *82*(3), 498.

Seery, N., Buckley, J., & Delahunty, T. (2015). Developing a Spatial Ability Framework to Support Spatial Ability Research in Engineering Education. In *The 6th Research in Engineering Education Symposium*. Retrieved from https://www.researchgate.net/profile/Jeffrey_Buckley/publication/308972809_Developing_a_Spatial_Ability_Framework_to_Support_Spatial_Ability_Research_in_Engineering_Education/links/57fc213208aea0db5a3f63a3.pdf

Shaw, E. J., Marini, J. P., Beard, J., Shmueli, D., Young, L., & Ng, H. (2016). *The Redesigned SAT® Pilot Predictive Validity Study: A First Look. Research Report 2016-1*. College Board. Retrieved from https://eric.ed.gov/?q=%22%22&ff1=souCollege+Board&id=ED563472

Shea, D. L., Lubinski, D., & Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology*, *93*(3), 604.

Shepard, R., & Cooper, L. (1982). Mental images and their transformations.

Silverman, I., Choi, J., & Peters, M. (2007). The hunter-gatherer theory of sex differences in spatial abilities: Data from 40 countries. *Archives of Sexual Behavior*, *36*(2), 261–268.

Smith, I. M. (1964). *Spatial ability: Its educational and social significance*. London: University of London Press Ltd.

Sorby, S. A. (2009). Educational Research in Developing 3-D Spatial Skills for Engineering Students. *International Journal of Science Education*, *31*(3), 459–480. https://doi.org/10.1080/09500690802595839

Sorby, S. A., & Baartmans, B. J. (1996). A Course for the Development of 3-D Spatial Visualization Skills. *Engineering Design Graphics Journal*, *60*(1), 13–20.

Sorby, S. A., & Baartmans, B. J. (2000). The Development and Assessment of a Course for Enhancing the 3-D Spatial Visualization Skills of First Year Engineering Students. *Journal of Engineering Education*, *89*(3), 301–307.

Sorby, S. A., Casey, B., Veurink, N., & Dulaney, A. (2013). The role of spatial training in improving spatial and calculus performance in engineering students. *Learning & Individual Differences*, *26*, 20–29. https://doi.org/10.1016/j.lindif.2013.03.010

Sorby, S. A., Leopold, C., & Gorska, R. (1999). Cross-cultural comparisons of gender differences in the spatial skills of engineering students. *Journal of Women and Minorities in Science and Engineering*, *5*(3).

Sorby, S. A., & Veurink, N. (2010a). Are The Visualization Skills Of First Year Engineering Students Changing? ASEE Conferences. Retrieved from https://peer.asee.org/16026

Sorby, S. A., & Veurink, N. (2010b). Long-term Results from Spatial Skills Intervention among First-Year Engineering Students. Presented at the Proceedings of the 65th Midyear Meeting of the Engineering Design Graphics Division of ASEE.

Spearman, C. (1904). “General Intelligence,” Objectively Determined and Measured. *The American Journal of Psychology*, *15*(2), 201–292. https://doi.org/10.2307/1412107

Spearman, C. (1927). *The abilities of man; their nature and measurement.* New York: Macmillan Co.

Sternberg, R. J. (2004a). *International handbook of intelligence*. Cambridge; New York: Cambridge University Press. Retrieved from http://dx.doi.org/10.1017/CBO9780511616648

Sternberg, R. J. (2004b). North American approaches to intelligence. In R. J. Sternberg (Ed.), *International handbook of intelligence* (pp. 411–444). Retrieved from https://books.google.com/books?hl=en&lr=&id=qbNLJl_L6MMC&oi=fnd&pg=PA411&dq=north+american+approaches+to+intelligence&ots=HIFKM_Urqa&sig=nB4yQt1SLrQox_b_2Eoo5abwDN8

Thornton, R., & Sokoloff, D. (1998). Assessing student learning of Newton’s laws: The Force and Motion Conceptual Evaluation and the Evaluation of Active Learning Laboratory and Lecture Curricula. *AMERICAN JOURNAL OF PHYSICS*, *66*(4), 338–352. https://doi.org/10.1119/1.18863

Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1951-00635-003&site=ehost-live

Uttal, D. H., & Cohen, C. A. (2012). *Spatial Thinking and STEM Education. When, Why, and How?* (Vol. 57). Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-84860615431&partnerID=40&md5=9e7bafd47ed3bb637fcef4311e09dfd7

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The Malleability of Spatial Skills: A Meta-Analysis of Training Studies. *Psychological Bulletin*, *139*(2), 352–402.

Van Gog, T., Paas, F., Van Merriëmboer, J. G., & Witte, P. (2005). Uncovering the Problem-Solving Process: Cued Retrospective Reporting Versus Concurrent and Retrospective Reporting. *Journal of Experimental Psychology. Applied*, *11*(4), 237–244. https://doi.org/10.1037/1076-898X.11.4.237

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, *47*(2), 599–604.

Veurink, N., & Sorby, S. A. (2011). Raising the Bar? Longitudinal Study to Determine Which Students Would Benefit Most From Spatial Training. Presented at the Proceedings of the ASEE 2011 Annual Conference and Exposition, Vancouver, BC, Canada.

Voyer, D. (2011). Time limits and gender differences on paper-and-pencil tests of mental rotation: A meta-analysis. *Psychonomic Bulletin & Review*, *18*(2), 267–277.

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, *101*(4), 817.

Walsh, L. N. (2009). A phenomenographic study of introductory physics students: approaches to problem solving and conceptualisation of knowledge. *Doctoral*, 61.

Walsh, L. N., Howard, R. G., & Bowe, B. (2007). Phenomenographic study of students' problem solving approaches in physics. *Physical Review Special Topics - Physics Education Research*, *3*(2), 020108. https://doi.org/10.1103/PhysRevSTPER.3.020108

Wise, J. C., Lee, S. H., Litzinger, T., Marra, R. M., & Palmer, B. (2004). A report on a four-year longitudinal study of intellectual development of engineering undergraduates. *Journal of Adult Development*, *11*(2), 103–110.

Yoon, S. Y. (2011). *Psychometric Properties of the Revised Purdue Spatial Visualization Tests: Visualization of Rotations (The Revised PSVT-R).* ERIC. Retrieved from http://eric.ed.gov/?id=ED534824

Zoltowski, C. B., Oakes, W. C., & Cardella, M. E. (2012). Students' Ways of Experiencing Human-Centered Design. *Journal of Engineering Education*, *101*(1), 28–59.

Appendix A Math pilot study problems set

Problem A1-1: 'Jug'

Stainless steel cylindrical jugs are made to hold a volume of 2 L. If the 1 litre mark is at 8.84 cm what is the radius of the jug to the nearest centimetre?

Problem A1-2: 'Pizza'

An 8 inch pizza is cut into 4 equal slices.

A 10 inch pizza is cut into 6 equal slices.

Which slices are larger?

Problem A1-3: 'Cleaner'

Brian, Peter and Janice are office cleaners. They are expected to dust, polish and vacuum the office. The time in minutes each takes to do each of the various jobs is shown in the table below.

	Dust	Vacuum	Polish
Brian	45	40	35
Janice	45	50	40
Peter	40	40	30

Which single job should their supervisor allocate to each cleaner so that the total time taken to clean the office is as short as possible?

Answer Choice (Grid)

Code	Dust	Vacuum	Polish
a	Janice	Brian	Peter
b	Peter	Brian	Janice
c	Brian	Janice	Peter
d	Peter	Janice	Brian

Problem A1-4: 'Sequence'

What are the next two terms in the following sequence?

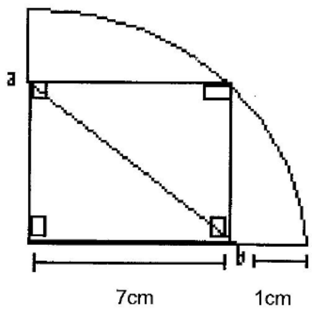
2, 7, 17, 37 ☐ ☐

Problem A2-1: 'Cans'

Drinks cans are made by stamping out circular discs from a sheet of metal.

The rectangular sheet from which the circles are stamped out measures 1 m by 2 m. If the cans have a radius of 10 cm, how many cans can be made from this sheet of metal?

Problem A2-2: 'Rectangle'



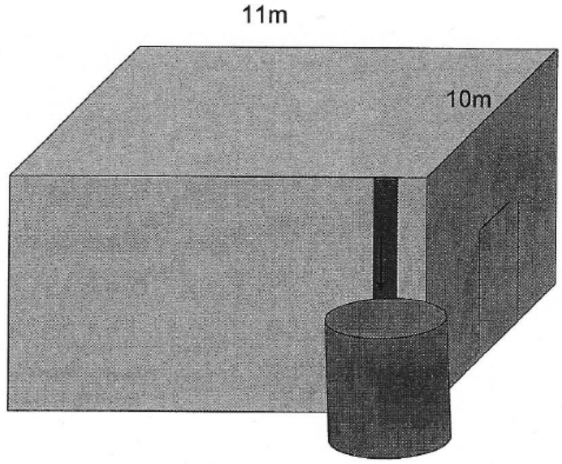
If a rectangle is drawn in the quadrant of a circle with the measurements indicated, what is the circumference of the circle in terms of pi?

Problem A2-3: 'Blood'

When blood samples are centrifuged the blood separates into two distinct layers one made up mainly of plasma the other made up of red blood cells.

A sample of blood was put in a flat bottomed test tube with a diameter of 3 cm. When the blood sample was added to the tube it filled the tube to a depth of 7.5cm. After centrifuging the red blood layer was 1.5 cm deep. What volume of plasma was in the sample?

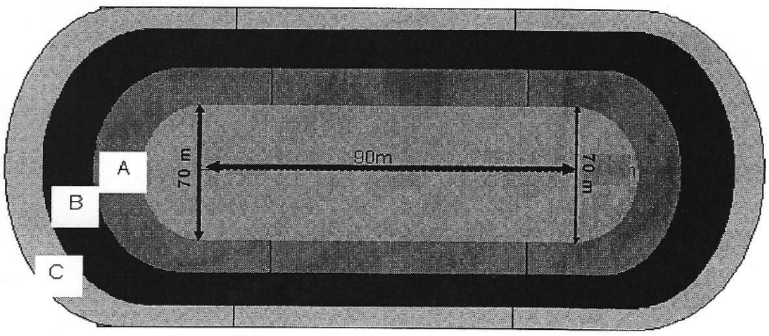
Problem A2-4: 'Rain'



The diagram above shows the dimensions of a flat roofed commercial shed. During one week 5 mm of rainfall fell on the roof of the shed. The rain was caught by gutters that flowed into a cylindrical water barrel with a diameter on 1 m. By how much did the depth of the water in the barrel increase (in cm) as a result of the rain? (Round your answer to the nearest cemtimetre)

Problem A3-1: 'Track'

Three athletes Ann, Bryan and Carol run around the running track as shown in the diagram below. Each of the athletes runs in a lane which is 1 metre wide.



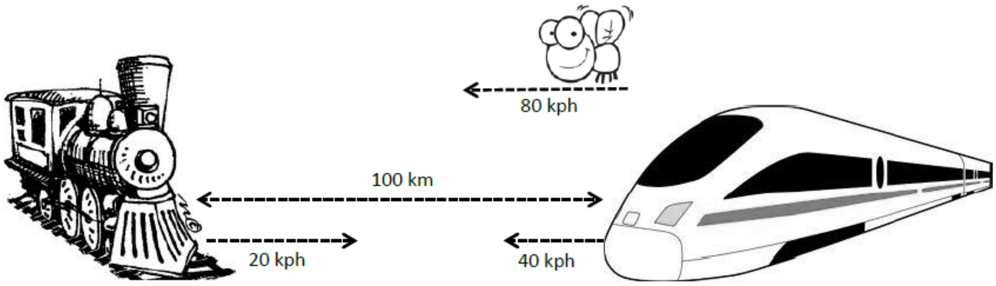
The organisers want the three athletes to compete in a 400m race. How much of a stagger must be given to Bryan and Carol to make the race fair? (i.e. how far ahead of Ann should they start?)

Problem A3-2: 'HCl'

You have been asked to provide 150 ml of hydrochloric acid at a 12% concentration. In the cabinet you find one bottle marked 10% HCl (10% hydrochloric acid) and another bottle marked 25% HCl (25% hydrochloric acid). How many ml of each should you use?

Problem A4-1: 'Fly'

The Crazy Zig-Zagging Fly



A fly at the front of the train travelling at 40 kph takes off flying at 80 kph until it reaches the front of the second train. It then immediately turns around and flies back towards the first train. It turns around again and flies back towards the other train.

As the trains get closer and closer together, the fly zig-zags back and forth over shorter and shorter distances, until the trains touch, and the fly is killed.

How far does the fly fly?

Problem A4-2: 'Egg drop'

You are in a 100 storey building and have 2 eggs. What is the minimum number of drops to determine the highest floor from which an egg can be dropped without breaking? (Both eggs can be broken).

Problem C2-2: 'Lawn'

A square lawn was extended in width by 2 m and in length by 3 m. The area of the new lawn is twice as big as the area of the old lawn. What are the measurements of the old lawn?

Problem B3-2: 'Jars & Pencils'

I have some pencils and some jars. If I put 4 pencils into each jar I will have one jar left over. If I put 3 pencils into each jar I will have one pencil left over.

How many pencils and how many jars are there?

Problem A4-3: 'Stadium'

A stadium can hold 25,000 people. People attending a regular event at the stadium must purchase a ticket in advance. When the ticket price is \$20, the expected attendance at an event is 12,000 people. The results of a survey carried out by the owners suggest that for every \$1 reduction, from \$20, in the ticket price, the expected attendance would increase by 1,000 people.

(i) Find the price at which tickets should be sold to give the maximum expected income and calculate this income.

(ii) Suppose that tickets are instead priced at a value that is expected to give a full attendance at the stadium. Find the difference between the income from the sale of tickets at this price and the maximum income calculated at (i) above.

Appendix B Math problems and questions

Problems

Problem 1: Lawn

A **square** lawn was extended in width by 2 m and in length by 3 m. The area of the new lawn is twice as big as the area of the old lawn. What are the dimensions of the old lawn?

(A lawn is an area of grass beside a house).

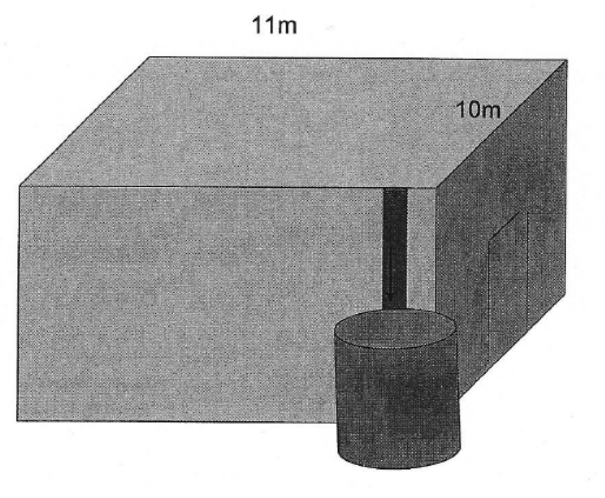
Problem 2. Jug

Stainless steel cylindrical jugs are made to hold a volume of 2 litres (2000 cm^3). If the 1 litre mark is at 8.84 cm what is the **radius** of the jug to the nearest centimetre?

Problem 3: Cans

Drink cans are made by stamping out circular discs from a sheet of metal. The rectangular sheet from which the discs are stamped out measures 1 m by 2 m. If the cans have a **radius** of 10 cm, how many discs can be made from this sheet of metal?

Problem 4: Rain



The diagram above shows the dimensions of a flat roofed commercial shed. During one week 5 mm of rain fell on the roof of the shed. The rain was collected by gutters that flowed into a cylindrical water barrel with a **diameter** of 1 m. By how much did the depth of the water in the barrel increase as a result of this rain?

Problem 5: Jars

I have some pencils and some jars. If I put 4 pencils into each jar I will have one jar left over. If I put 3 pencils into each jar I will have one pencil left over.

How many pencils and how many jars are there?

Problem 6: Blood

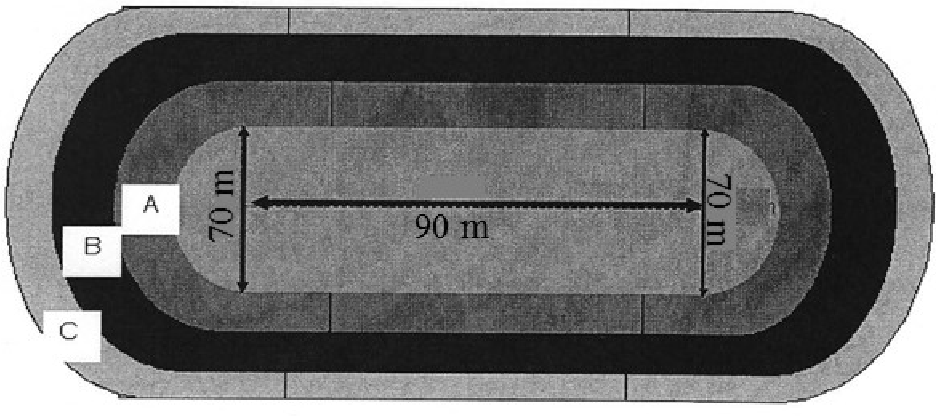
When blood samples are centrifuged the blood separates into two distinct layers – one made up mainly of plasma and the other made up of red blood cells.

A sample of blood was put in a flat bottomed test tube with a **diameter** of 3 cm. When the blood sample was added to the tube it filled the tube to a depth of 7.5 cm. After centrifuging, the red blood layer was 1.5 cm deep. What volume of plasma was in the sample?

Problem 7: Track

Three athletes – Ann, Bryan and Carol – run around the running track as shown in the diagram below. Each of the athletes runs in a lane which is 1 metre wide.

The organizers want the three athletes to compete in a 400 m race. How much of a stagger must be given to Bryan and Carol to make the race fair, i.e. how far ahead of Ann should they start?



Problem 8: HCl

You have been asked to provide 150 ml of hydrochloric acid at a 12% concentration. In the cabinet you find one bottle marked 10% HCl (10% hydrochloric acid) and another bottle marked 25% HCl (25% hydrochloric acid). How many ml of each should you use?

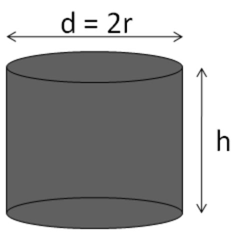
Questions

Question 1

Find the roots of $2x^2 + 6x - 8 = 0$ using factoring.

Question 2

What is the volume of this cylinder?



Question 3

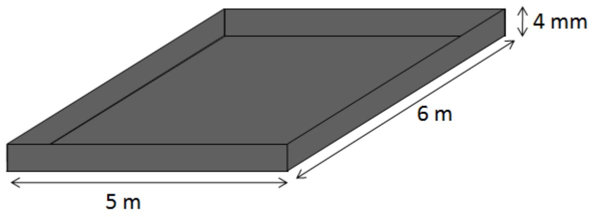
How many centimetres are in a metre?

Question 4

What is the equation for the area of a circle?

Question 5

What is the volume of this tank?



Question 6

Determine the value of x and y by solving these two equations

$x + y = 6$

$-3x + y = 2$

Appendix C DIRECT Test

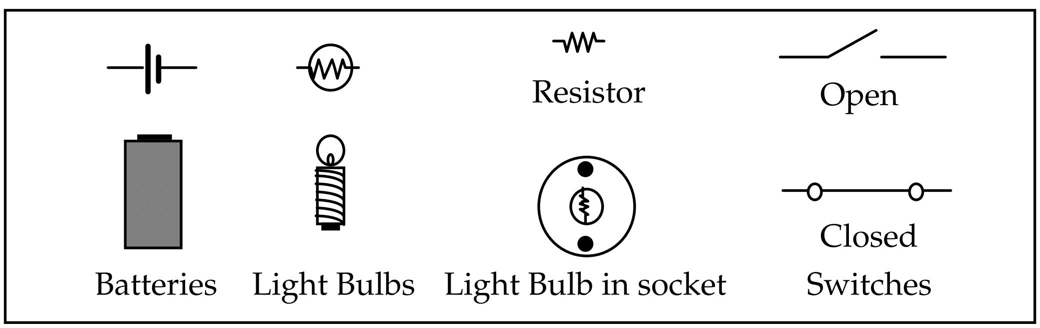
Diagnostic Electric Circuits Test

Instructions

Wait until you are told to begin, then turn to the next page and begin working. Answer each question as accurately as you can. There is only one correct answer for each item. Feel free to use a calculator and scratch paper if you wish. You will have approximately half an hour to complete the test. If you finish early, check your work before handing in both the answer sheet and the test booklet.

Additional comments about the test

All light bulbs, resistors, and batteries should be considered identical unless you are told otherwise. The battery is to be assumed ideal, that is to say, the internal resistance of the battery is negligible. In addition, assume the wires have negligible resistance. Below is a key to the symbols used on this test. Study them carefully before you begin the test.



1) Are charges used up in the production of light in a light bulb?

(A) Yes, charge is used up. Charges moving through the filament produce "friction" which heats up the filament and produces light.

(B) Yes, charge is used up. Charges are emitted as photons and are lost.

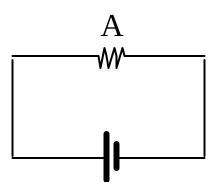
(C) Yes, charge is used up. Charges are absorbed by the filament and are lost.

(D) No, charge is conserved. Charges are simply converted to another form such as heat and light.

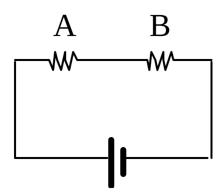
(E) No, charge is conserved. Charges moving through the filament produce "friction" which heats up the filament and produces light.

2) How does the power delivered to resistor A change when resistor B is added to the circuit? The power delivered to resistor A _____.

(A) Quadruples (4 times)
(B) Doubles
(C) Stays the same
(D) Reduces by half
(E) Reduces by one quarter (1/4)



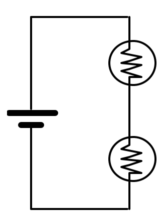
Before

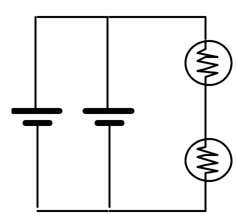


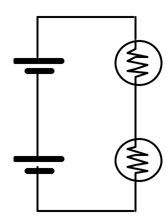
After

3) Which circuit or circuits have the GREATEST energy delivered to them per second?

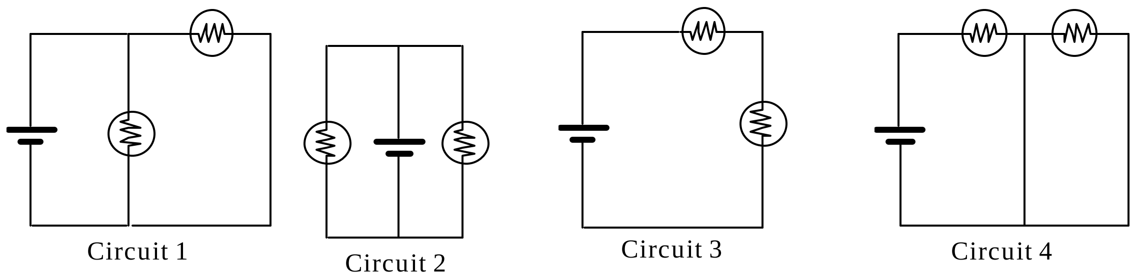
(A) Circuit 1
(B) Circuit 2
(C) Circuit 3
(D) Circuit 1 = Circuit 2
(E) Circuit 2 = Circuit 3


Circuit 1


Circuit 2


Circuit 3

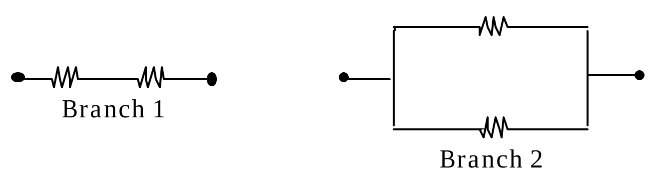
4) Which circuit or circuits below represent a circuit consisting of two light bulbs in parallel with a battery?



(A) Circuit 1
(B) Circuit 2
(C) Circuit 3
(D) Circuits 1 and 2
(E) Circuits 1, 2, and 4

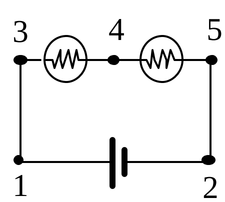
5) Compare the resistance of branch 1 with that of branch 2. A branch is a section of a circuit. The resistance of branch 1 is _____ branch 2.

(A) Four times
(B) Double
(C) The same as
(D) Half
(E) One quarter (1/4)



6) Rank the potential difference between points 1 and 2, points 3 and 4, and points 4 and 5 in the circuit shown below from HIGHEST to LOWEST.

(A) 1 and 2; 3 and 4; 4 and 5
(B) 1 and 2; 4 and 5; 3 and 4
(C) 3 and 4; 4 and 5; 1 and 2
(D) 3 and 4 = 4 and 5; 1 and 2
(E) 1 and 2; 3 and 4 = 4 and 5



7) Compare the brightness of the bulb in circuit 1 with that in circuit 2. Which bulb is BRIGHTER?

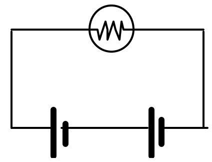
(A) Bulb in circuit 1 because two batteries in series provide less voltage

(B) Bulb in circuit 1 because two batteries in series provide more voltage

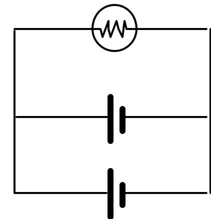
(C) Bulb in circuit 2 because two batteries in parallel provide less voltage

(D) Bulb in circuit 2 because two batteries in parallel provide more voltage

(E) Neither, they are the same



Circuit 1



Circuit 2

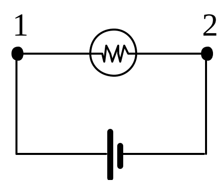
8) Compare the current at point 1 with the current at point 2. Which point has the LARGER current?

(A) Point 1

(B) Point 2

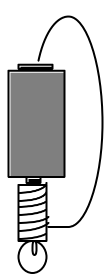
(C) Neither, they are the same. Current travels in one direction around the circuit.

(D) Neither, they are the same. Currents travel in two directions around the circuit.

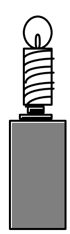


9) Which circuit or circuits will light the bulb?

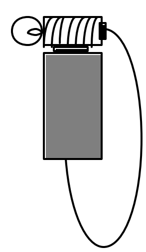
(A) Circuit 1
(B) Circuit 2
(C) Circuit 3
(D) Circuits 1 and 3
(E) Circuits 1, 3, and 4



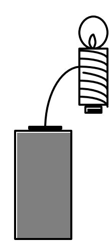
Circuit 1



Circuit 2



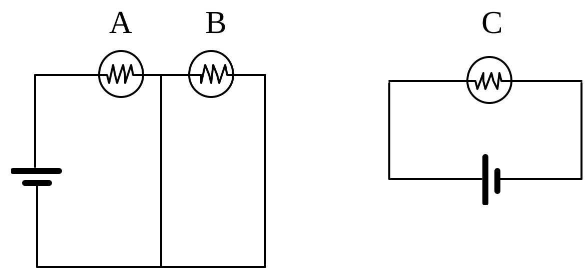
Circuit 3



Circuit 4

10) Compare the brightness of bulbs A, B, and C in these circuits. Which bulb or bulbs are the BRIGHTEST?

(A) A
(B) B
(C) C
(D) A = B
(E) A = C



11) Why do the lights in your home come on almost instantaneously when you turn on the switch?

(A) When the circuit is completed, there is a rapid rearrangement of surface charges in the circuit.

(B) Charges store energy. When the circuit is completed, the energy is released.

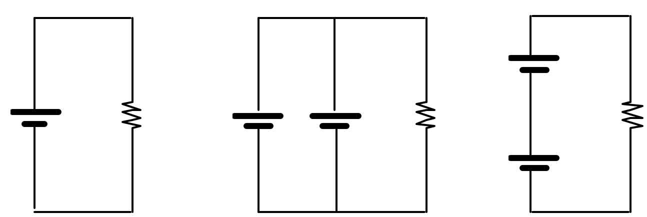
(C) Charges in the wire travel very fast.

(D) The circuits in a home are wired in parallel. Thus, a current is already flowing.

(E) Charges in the wire are like marbles in a tube. When the circuit is completed, the charges push each other through the wire.

12) Consider the power delivered to each of the resistors shown in the circuits below. Which circuit or circuits have the LEAST power delivered to them?

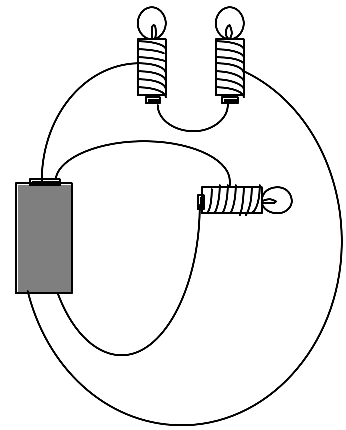
(A) Circuit 1
(B) Circuit 2
(C) Circuit 3
(D) Circuit 1 = Circuit 2
(E) Circuit 1 = Circuit 3

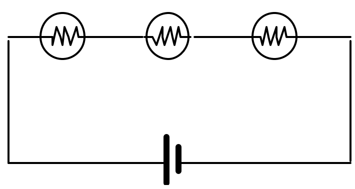


Circuit 1 Circuit 2 Circuit 3

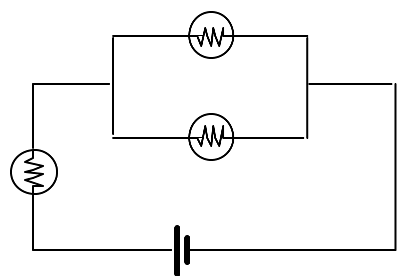
13) Which schematic diagram best represents the realistic circuit shown below?

(A) Circuit 1
(B) Circuit 2
(C) Circuit 3
(D) Circuit 4
(E) None of the above

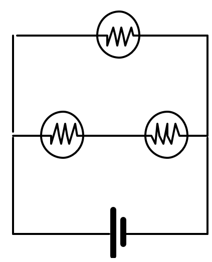




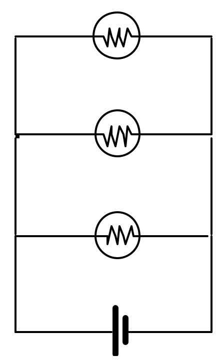
Circuit 1



Circuit 2



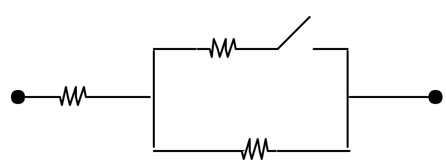
Circuit 3



Circuit 4

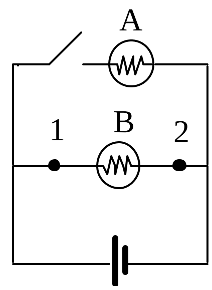
14) How does the resistance between the endpoints change when the switch is closed?

(A) Increases by R
(B) Increases by R/2
(C) Stays the same
(D) Decreases by R/2
(E) Decreases by R



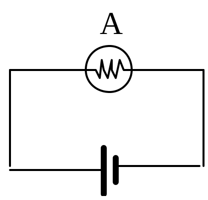
15) What happens to the potential difference between points 1 and 2 when the switch is closed?

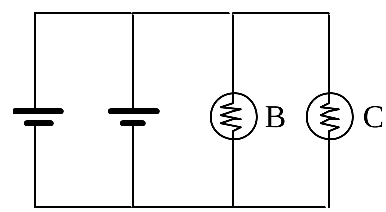
(A) Quadruples (4 times)
(B) Doubles
(C) Stays the same
(D) Reduces by half
(E) Reduces by one quarter (1/4)



16) Compare the brightness of bulb A with bulb B. Bulb A is _____ bright as Bulb B.

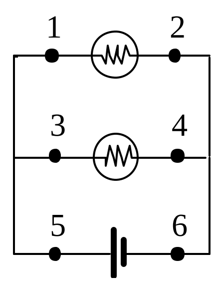
(A) Four times as
(B) Twice as
(C) Equally
(D) Half as
(E) One fourth (1/4) as



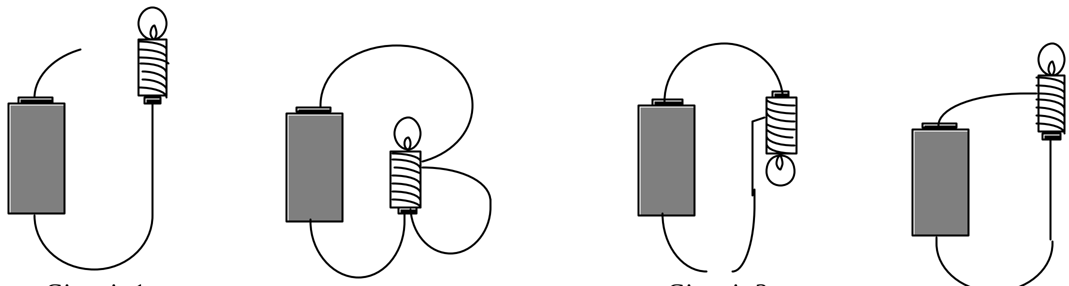


17) Rank the currents at points 1, 2, 3, 4, 5, and 6 from HIGHEST to LOWEST.

(A) 5, 3, 1, 2, 4, 6
(B) 5, 3, 1, 4, 2, 6
(C) 5 =6, 3 = 4, 1 =2
(D) 5 = 6, 1 = 2 = 3 = 4
(E) 1 = 2 = 3 = 4 = 5 = 6



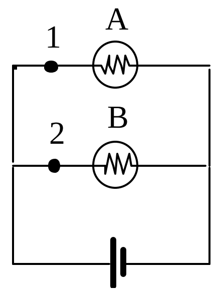
18) Which circuit(s) will light the bulb?



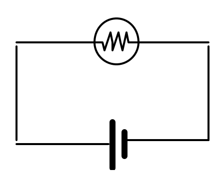
(A) Circuit 1
(B) Circuit 2
(C) Circuit 4
(D) Circuits 2 and 4
(E) Circuits 1 and 3

19) What happens to the brightness of bulbs A and B when a wire is connected between points 1 and 2?

(A) Both increase
(B) Both decrease
(C) They stay the same
(D) A becomes brighter than B
(E) Neither bulb will light



20) Is the electric field zero or non-zero inside the bulb filament?



(A) Zero because the filament is a conductor.

(B) Zero because a current is flowing.

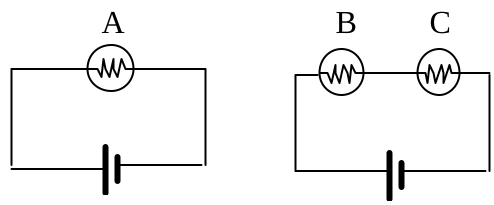
(C) Zero because there are charges on the surface of the filament.

(D) Non-zero because a current is flowing which produces the field.

(E) Non-zero because there are charges on the surface of the filament which produce the field.

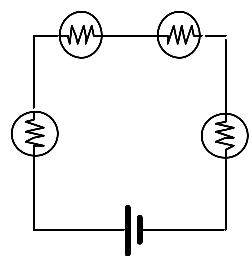
21) Compare the energy delivered per second to each light bulb shown below. Which bulb or bulbs have the LEAST energy delivered to them per second?

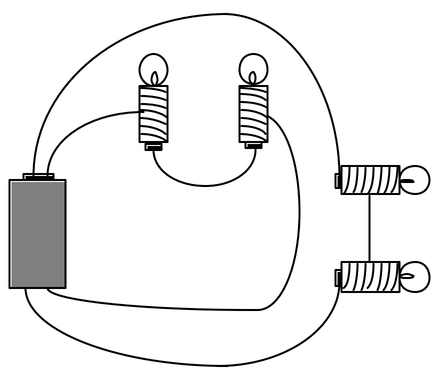
(A) A
(B) B
(C) C
(D) B = C
(E) A = B = C

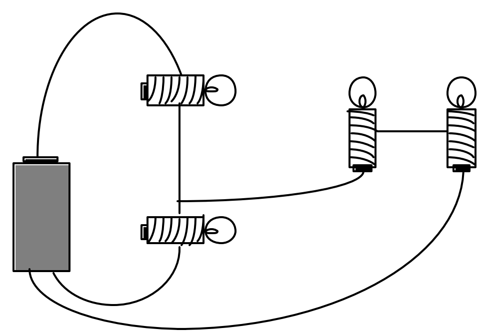


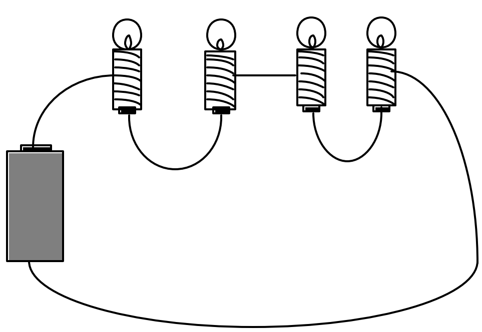
22) Which realistic circuit or circuits represent the schematic diagram shown below?

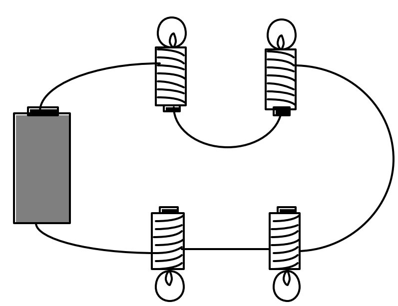
(A) Circuit 2
(B) Circuit 3
(C) Circuit 4
(D) Circuits 1 and 2
(E) Circuits 3 and 4




Circuit 1

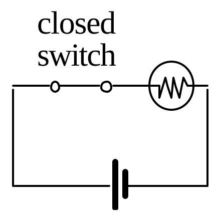

Circuit 2


Circuit 3


Circuit 4

23) Immediately after the switch is opened, what happens to the resistance of the bulb?

(A) The resistance goes to infinity.
(B) The resistance increases.
(C) The resistance decreases.
(D) The resistance stays the same.
(E) The resistance goes to zero.

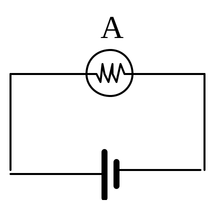


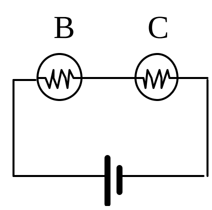
24) If you double the current through a battery, is the potential difference across a battery doubled?

(A) Yes, because Ohm's law says $V = IR$.
(B) Yes, because as you increase the resistance, you increase the potential difference.
(C) No, because as you double the current, you reduce the potential difference by half.
(D) No, because the potential difference is a property of the battery.
(E) No, because the potential difference is a property of everything in the circuit.

25) Compare the brightness of bulb A with bulb B. Bulb A is _____ bright as bulb B.

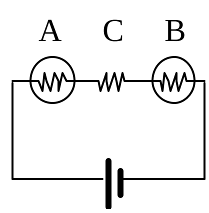
(A) Four times as
(B) Twice as
(C) Equally
(D) Half as
(E) One fourth (1/4) as



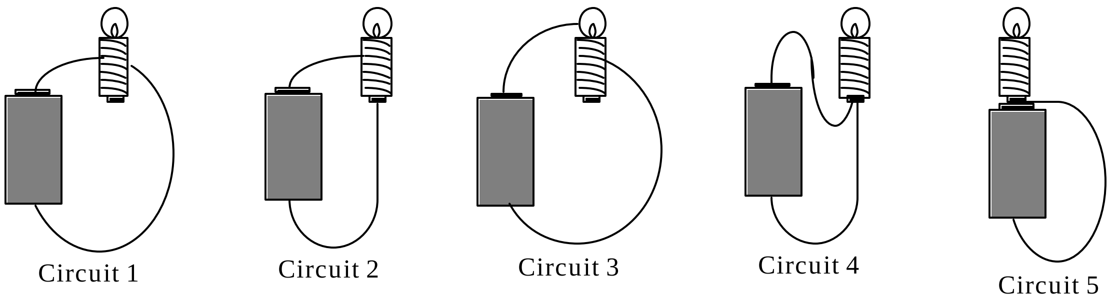


26) If you increase the resistance C, what happens to the brightness of bulbs A and B?

(A) A stays the same, B dims
(B) A dims, B stays the same
(C) A and B increase
(D) A and B decrease
(E) A and B remain the same



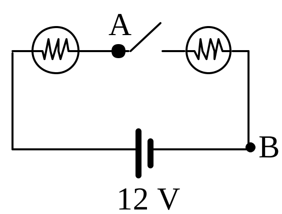
27) Will all the bulbs be the same brightness?



(A) Yes, because they all have the same type of circuit wiring.
(B) No, because only Circuit 2 will light.
(C) No, because only Circuits 4 and 5 will light.
(D) No, because only Circuits 1 and 4 will light.
(E) No, Circuit 3 will not light but Circuits 1, 2, 4, and 5 will.

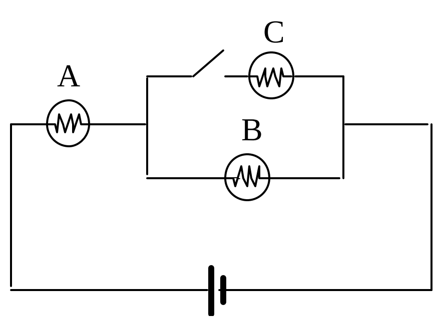
28) What is the potential difference between points A and B?

(A) 0 V
(B) 3 V
(C) 6 V
(D) 12 V
(E) None of the above



29) What happens to the brightness of bulbs A and B when the switch is closed?

(A) A stays the same, B dims
(B) A brighter, B dims
(C) A and B increase
(D) A and B decrease
(E) A and B remain the same



Appendix D MPT example

MPT D Sample Problems for Spatial Skills Research

(1) Which is a possible solution to $\frac{7}{x-9}+\frac{x+9}{7}=\frac{79}{97}$ is...
(**mostly routine**)

(2) Rewrite A in terms of $\log 7$ and $\log 3$, where $A=\log 189$.
(**less routine, requires more problem solving**)

(3) Presented and similar math concepts as Practice Test Problem 1
(*Uses function notation instead of numerical values*)
(**mostly routine**)

(4) Presented and similar math concepts as Practice Test Problem 4c
(**mostly routine**)

(5) Which of the following numbers may be multiplied by $7+9i$ to obtain a real number?
(**mostly routine**)

(6) Presented and similar math concepts as Practice Test Problem 1
(*Uses numerical values like practice problem*)
(**mostly routine**)

(7) Presented and similar math concepts as Practice Test Problem 1
(*Uses function notation instead of numerical values*)
(**mostly routine**)

(8) Let x vary inversely as y. When x is 21, y is 4. When x is 7, y is ...
(**mostly routine**)

(9) Presented and similar math concepts as Practice Test Problem 2f
(**mostly routine**)

(10) If $f(x)=e^x$ and if $a,b\in\mathbb{Z}$, then $f(a)\cdot f(b)$ can be written as ...
(**less routine, requires more problem solving**)

(11) How many pounds of M&Ms that cost $1.60 per pound must be mixed with 7 pounds of Smarties that cost $2.20 per pound to make a mixture of sweets that costs $2.00 per pound?
(**less routine, requires more problem solving**)

(12) Presented and similar math concepts as Practice Test Problem 2d
(**mostly routine**)

(13) Presented and similar math concepts as Practice Test Problem 14b
(**mostly routine**)

(14) Similar math concepts as Practice Test Problem 7. Presented as "which of the following are not"...
(**mostly routine**)

(15) What is the velocity in radians per minute of a Ferris wheel that makes 60 revolutions an hour?
(**less routine, requires more problem solving**)

(16) Presented and similar math concepts as Practice Test Problem 8
(*Uses radians*)
(**mostly routine**)

(17) The function $f(x) = \cot x$ is not defined for...
(*Students need to use the definition. There are no calculations*)
(**less routine, requires more problem solving**)

(18) Presented and similar math concepts as Practice Test Problem 8
(*Uses degrees*)
(**mostly routine**)

(19) What is the exact value of $\sin(75°)$.
(*Students would need to use a series of trigonometric identities. Calculators could not be used as an exact answer such as* $\frac{\sqrt{6}+1}{4}$ *is the answer*)
(**less routine, requires more problem solving**)

(20) If $f(x) = \sin x$, then $f(\frac{\pi}{2} - x)$ is equal to...
(*Students need to use the definition. There are no calculations*)
(**less routine, requires more problem solving**)

(21) In a triangle, sides a, b and c are opposite angles A, B and C, respectively. If $a = 3$, $b = 4$ and $c = 5$, find a formula for A.
(**mostly routine**)

(22) Vertices A, B, C, D and E of pentagon $ABCDE$ have the following coordinates $A(2, -2)$, $B(-4, 2)$, $C(-1, 6)$, $D(4, 8)$ and $E(2, 6)$. The sine of angle CDE is...
(**less routine, requires more problem solving**)

(*If I were to guess, I would assume more students get 21 correct than 22. They use the same mathematical concepts, however 22 is asked differently and would require a higher level of thinking*)

(23) Presented and similar math concepts as Practice Test Problem 12
(*Students need to use the definition. There are no calculations*)
(**less routine, requires more problem solving**)

(24) In a parallelogram, an acute angle θ is created by side length a and b. Which is a formula for the area of the parallelogram?
(*Students need to use the definition. There are no calculations*)
(**less routine, requires more problem solving**)

(25) How many number of real solutions between 0 and 2π exist for the equation $\sin^2(x) - \frac{1}{2} = 0$?
(**less routine, requires more problem solving**)

(*This problem is interesting. Normally students are asked to simply solve for all solutions. However, when asked how many they sometimes struggle.*
Also, this is a way to prevent guessing. If we give the solutions, they can plug them in. However, there is no way to plug-and-chug if we ask in this form.)

Math Placement D-Test Sample Problems

(1) Simplify as much as possible:

$$\frac{\sqrt[3]{\frac{a^9}{4}}}{\sqrt[3]{16b^3}}$$

(2) Solve the following equations and inequalities for x.

(a) $\log_x 81 = 4$

(b) $16^{x-1} = 8$

(c) $\left|x^2 - 26\right| = 10$

(d) $|-3x + 1| < 2$

(e) $x + \sqrt{2x + 6} = 9$

(f) $(2x - 3)(3x + 2) \leq 0$

(g) $\dfrac{2x - 1}{x - 2} \leq 2$

(3) Find the distance between the two points (3,4) and (-5,8).

(4) Let $f(x) = \sqrt{1 + x}$ and $g(x) = \dfrac{3x^2}{x^2 + 1}$.

(a) Find $g[f(x)]$

(b) What is the range of g?

(c) Does f^{-1} exist? Does g^{-1} exist? If possible give f^{-1} and $g^{-}1$ and specify their domains.

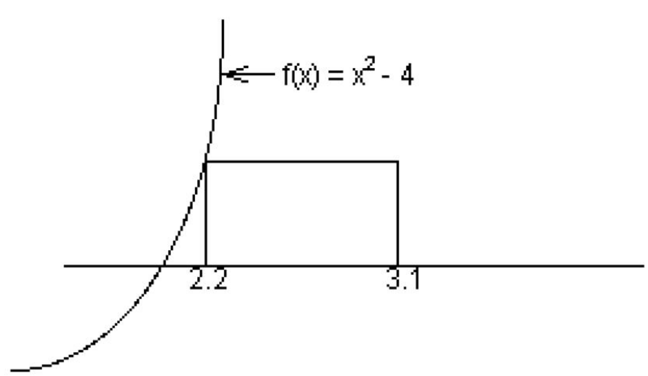
(5) Sketch the graphs of the following equations.

(a) $x^2 + 9y^2 = 81$

(b) $y = \log_2 8x$

(c) $y = x^2 + 4x + 1$ (label vertex)

(6) Find the area of the rectangle pictured.



(7) Give the Center and the radius of the circle $x^2 + y^2 - 6x + 8y = 0$.

(8) If $f(x) = \sin(2x)$, what is $f(\pi/4)$?

(9) What is $\sin\theta$ if θ is in standard position and $(2, -7)$ is on its terminal side?

(10) Find $\cos\left(2\sin^{-1}\dfrac{\sqrt{2}}{2}\right)$.

(11) Graph these functions. Label your graphs carefully.

(a) $y = \sin(2x)$, $0 \le x \le 2\pi$

(b) $y = \cos^{-1}(x)$ or $y = \arccos(x)$

(c) $y = x\cos(x)$, $-\pi \le x \le \pi$

(12) $\sin\theta\tan\theta + \cos\theta$ is one of the trigonometric functions. Which one?

(13) A regular hexagon (6 sides) is inscribed in a circle of radius 10 feet. Find its perimeter.

(14) If $z_1 = 2 - 3i$ and $z_2 = 3 + i$, find

(a) $|z_1|$

(b) $z_1 z_2$

(c) $\dfrac{z_1}{z_2}$

(15) For what r and θ does $r(\cos\theta + i\sin\theta) = 8i$?

(16) If $z = (\cos 30° + i\sin 30°)$, find a value of n, $n \neq 0$, for which $z^n = 1$.

Appendix E Ethics documents

DIT site ethics approval

Institiúid Teicneolaíochta Átha Cliath, Sráid Caoimhín, Baile Átha Cliath 8, Éire
Dublin Institute of Technology, Kevin Street, Dublin 8, Ireland
www.dit.ie/graduateresearchschool

SCOIL TAIGHDE IARCHÉIME / GRADUATE RESEARCH SCHOOL
Professor Mary McNamara

5th November 2015

Brian Bowe

Re: Ethical Clearance Ref 15-42

Dear Brian,

I am pleased to inform you that the following project:

'The impact of students' spatial visualization skills on their success and retention in science and engineering disciplines'

which you submitted to the Research Ethical Committee has been approved. The committee would like to wish you very best of luck with the rest of research project. If you have any further queries, please do not hesitate to contact Aisling Heyenga on (01) 402 7920 or at aisling.heyenga@dit.ie.

Yours sincerely

Aisling Heyenga

Dublin Institute of Technology
Research Ethics Committee

Informed consent 1 OSU Jan 2016

Department of Electrical Engineering

Title of Research:

Spatial Visualization and Engineering Problem Solving

Purpose of Research:

This study will investigate how spatial reasoning skills as measured by the mental cutting test (MCT) relate to engineer problem solving. Your participation in this study will help us to further understand the cognitive links between spatial skills and engineering problem solving. This study is intended for undergraduate students who are enrolled in an undergraduate degree program at the University of Nebraska-Lincoln (UNL) or the Ohio State University (OSU). You must be 19 years of age or older in order to participate in this research.

Procedures:

This study has two phases, Phase 1 and Phase 2.

Participation in Phase 1 of this study will require approximately 30 minutes of your time. This time will be spent taking the mental cutting test. If you are a UNL student, participation will take place in 329 Scott Engineering Center or a similar private room. If you are an OSU student, participation will take place in 224 Hitchcock Hall or a similar private room.

Phase 1 of this study involves you taking the written or electronic version of the mental cutting test, a well-known and validated measure of spatial reasoning skills. Neither your name nor any other identifying information will be associated with your score on this mental cutting test. Only the research team will know the identity of students taking the test. The data from the scores on the mental cutting test will only be publically reported in aggregated form.

Please note that the research team has investigators from both UNL and OSU and all members of the research team will have access to the data. Data will be shared across institutions using secure electronic mechanisms and all members of the research team will protect the confidentiality of all study participants.

You **may** be asked to participate in a second phase of this study. Participation in Phase 1 of the study does not require you to participate in phase 2 of the study. However, by participating in Phase 1, you agree to be contacted about Phase 2 of the study unless you have already withdrawn from the study.

Phase 2 of the study has a separate informed consent form.

Benefits:

There are no known direct benefits associated with participation in this research.

Risks and/or Discomforts:

There are no known risks or discomforts associated with participation in this research.

Confidentiality:

Any information obtained during this study that could identify you will be kept strictly confidential. The data will be stored in a locked cabinet in the investigator's office and will only be seen by the investigator during the study and for five (5) years after the study is complete. The aggregated data will also be stored on a

209N Scott Engineering Center / Lincoln, NE 68588-0511
(402) 472-3771 / FAX (402) 472-4732

password protected computer. The information obtained in this study may be published in scientific journals or presented at scientific meetings but the data will be reported as aggregated data.

Compensation:
You will receive a $15 electronic gift card for participating in Phase 1 of this research.

Opportunity to Ask Questions:

You may ask any questions concerning this research at anytime by contacting Dr. Lance C. Pérez at 402-472-6258 or lperez@unl.edu or Dr. Sheryl Sorby at 614-247-8953 or sorby.1@osu.edu. If you would like to speak to someone else, please call the UNL Research Compliance Services Office at 402-472-6965 or irb@unl.edu.

Freedom to Withdraw:

Participation in this study is voluntary. You can refuse to participate or withdraw at any time without harming your relationship with the researchers or the University of Nebraska-Lincoln, [other organization] or in any other way receive a penalty or loss of benefits to which you are otherwise entitled.

Consent, Right to Receive a Copy:

You are voluntarily making a decision whether or not to participate in this research study. Your signature certifies that you have decided to participate having read and understood the information presented. You will be given a copy of this consent form to keep.

Signature of Participant:

____ ______________________________

Signature of Research Participant *Date*

Name and Phone number of investigator(s)

Lance C. Pérez, PhD
402-472-6258
lperez@unl.edu

Sheryl Sorby, PhD
614-247-8953
sorby.1@osu.edu

Informed consent 2 OSU Jan 2016

Department of Electrical Engineering

Title of Research:

Spatial Visualization and Engineering Problem Solving

Purpose of Research:

This study will investigate how spatial reasoning skills as measured by the mental cutting test (MCT) relate to engineer problem solving. Your participation in this study will help us to further understand the cognitive links between spatial skills and engineering problem solving. This study is intended for undergraduate students who are enrolled in an undergraduate degree program at the University of Nebraska-Lincoln (UNL) or the Ohio State University (OSU). You must be 19 years of age or older in order to participate in this research.

Procedures:

This study has two phases, Phase 1 and Phase 2. You have already completed Phase 1 of the study and are now being asked to participate in Phase 2.

Participation in Phase 2 of this study will require approximately 2 hours of your time. If you are a UNL student, participation will take place in 329 Scott Engineering Center, C89 East Stadium or a similar private room. If you are an OSU student, participation will take place in 224 Hitchcock Hall or a similar private room.

Phase 2 of this study involves you attempting to solve four written problems that are representative of problems you would encounter in an undergraduate engineering degree program. The work you do to solve these problems will be recorded using a LiveScribe pen that you will use to write on paper. You will attempt to solve these problems while being monitored with a noninvasive electroencephalogram (EEG) sensor. Electroencephalography is a non-invasive method to record electrical activity of the brain along the scalp. The EEG monitor used in this research consists of a headset with several sensors that is placed on the your head, much like a hat, and does not require any modifications to the subject.

Neither your name nor any other identifying information will be associated with any data collected during this study. Only the research team will know the identity of students participating. The data from the study will only be publically reported in forms that do not identify the participants.

Please note that the research team has investigators from both UNL and OSU and all members of the research team will have access to the data. Data will be shared across institutions using secure electronic mechanisms and all members of the research team will protect the confidentiality of all study participants.

Benefits:

There are no known direct benefits associated with participation in this research.

Risks and/or Discomforts:

There are no known risks or discomforts associated with participation in this research.

Confidentiality:

Any information obtained during this study that could identify you will be kept strictly confidential. The data will be stored in a locked cabinet in the investigator's office and will only be seen by the investigator during the study and for five (5) years after the study is complete. The aggregated data will also be stored on a

209N Scott Engineering Center / Lincoln, NE 68588-0511
(402) 472-3771 / FAX (402) 472-4732

password protected computer. The information obtained in this study may be published in scientific journals or presented at scientific meetings but the data will be reported as aggregated data.

Compensation:
You will receive a $100 check for completing the protocol in Phase 2 of this research. If you choose to withdraw from the study after more than 1 hour, but before completing the phase 2 protocol, you will receive a check for $50. If you choose to withdraw from the study less than 1 hour into the protocol, you will not receive any compensation.

All participants should note that this compensation is taxable income and you will be required to sign a receipt when you receive the compensation.

For participants from UNL, payments will be processed by the UNL Office of the Bursar and you will be required to provide your social security number in order to receive compensation.

For participants from OSU, you will receive a check written from an Ohio State University Research Foundation account upon completion of your participation.

Opportunity to Ask Questions:

You may ask any questions concerning this research at anytime by contacting Dr. Lance C. Pérez at 402-472-6258 or lperez@unl.edu or Dr. Sheryl Sorby at 614-247-8953 or sorby.1@osu.edu. If you would like to speak to someone else, please call the UNL Research Compliance Services Office at 402-472-6965 or irb@unl.edu.

Freedom to Withdraw:

Participation in this study is voluntary. You can refuse to participate or withdraw at any time without harming your relationship with the researchers or the University of Nebraska-Lincoln, [other organization] or in any other way receive a penalty or loss of benefits to which you are otherwise entitled.

Consent, Right to Receive a Copy:

You are voluntarily making a decision whether or not to participate in this research study. Your signature certifies that you have decided to participate having read and understood the information presented. You will be given a copy of this consent form to keep.

Signature of Participant:

____ _____________________________
Signature of Research Participant *Date*

Name and Phone number of investigator(s)

Lance C. Pérez, PhD
402-472-6258
lperez@unl.edu

Sheryl Sorby, PhD
614-247-8953
sorby.1@osu.edu

Informed consent 1 OSU August 2016

Consent Form

The Ohio State University Consent to Participate in Research

Study Title:	**Spatial Visualization and Engineering Problem-Solving.**
Researcher:	Sheryl A. Sorby
Sponsor:	National Science Foundation

This is a consent form for research participation. It contains important information about this study and what to expect if you decide to participate.

Please consider the information carefully. Feel free to ask questions before making your decision whether or not to participate. Your participation in the research project is voluntary. Your participation (or non-participation) will not impact your grade in any way at The Ohio State University.

Purpose:

This study will investigate how spatial reasoning skills as measured by the mental cutting test (MCT) relate to engineering problem solving. Your participation in this study will help us to further understand the cognitive links between spatial skills and engineering problem solving. This study is intended for undergraduate students who are enrolled in an undergraduate engineering degree program at the Ohio State University (OSU). You must be 18 years of age or older in order to participate in this research.

Procedures/Tasks:

This study has two phases, Phase 1 and Phase 2. Phase 1 will occur at the beginning of fall semester and Phase 2 will occur at the end of fall semester. Each phase will take approximately 1.5 hours of your time for a total of 3 hours overall. Part 1 of each phase consists of your completion of the Mental Cutting Test, a test of spatial cognition. This part will take approximately 20-25 minutes.

Participation in part 2 of each phase will require approximately 1-1.25 hours of your time.

Part 2 of this study involves you attempting to solve several written problems that are representative of problems you would encounter in an undergraduate engineering degree program. The work you do to solve these problems will be recorded using a LiveScribe pen that you will use to write on paper. You will attempt to solve these problems while being monitored with a noninvasive electroencephalogram (EEG) sensor. Electroencephalography is a non-invasive method to record electrical activity of the brain along the scalp. The EEG monitor used in this research consists of a headset with several sensors that is placed on the your head, much like a hat, and does not require any modifications to you.

Neither your name nor any other identifying information will be associated with any data collected during this study. Only the research team will know the identity of students participating. The data from the study will only be publically reported in forms that do not identify the participants.

At the end of each phase, you will receive an electronic gift card from Amazon for $75 (Phase 1) and $100 (Phase 2).

Duration:

Each session should take approximately 1.5 hours to complete. You will spend a total of 3 hours in two sessions.

You may leave the study at any time. If you decide to stop participating in the study, there will be no penalty to you, and you will not lose any benefits to which you are otherwise entitled. Your decision will not affect your future relationship with The Ohio State University.

Risks and Benefits:

There is minimal risk involved in this study. Your data may inadvertently be released; however, since the data on this test does not reflect your grade in any course in any way, the risk that confidential information will be accessed is extremely low. The benefit to your participation in this study is that this will inform future efforts in improving spatial skills for engineering students.

Confidentiality:

Efforts will be made to keep your study-related information confidential. However, there may be circumstances where this information must be released. For example, personal information regarding your participation in this study may be disclosed if required by state law. Also, your records may be reviewed by the following groups (as applicable to the research):

- Office for Human Research Protections or other federal, state, or international regulatory agencies;
- The Ohio State University Institutional Review Board or Office of Responsible Research Practices;

- The sponsor, if any, or agency (including the Food and Drug Administration for FDA-regulated research) supporting the study.

Participant Rights:

You may refuse to participate in this study without penalty or loss of benefits to which you are otherwise entitled. If you are a student or employee at Ohio State, your decision will not affect your grades or employment status.

If you choose to participate in the study, you may discontinue participation at any time without penalty or loss of benefits. By signing this form, you do not give up any personal legal rights you may have as a participant in this study.

An Institutional Review Board responsible for human subjects research at The Ohio State University reviewed this research project and found it to be acceptable, according to applicable state and federal regulations and University policies designed to protect the rights and welfare of participants in research.

Contacts and Questions:
For questions, concerns, or complaints about the study, or you feel you have been harmed as a result of study participation, you may contact Dr. Sheryl Sorby: sorby.1@osu.edu.

For questions about your rights as a participant in this study or to discuss other study-related concerns or complaints with someone who is not part of the research team, you may contact Ms. Sandra Meadows in the Office of Responsible Research Practices at 1-800-678-6251.

Signing the consent form

I have read (or someone has read to me) this form and I am aware that I am being asked to participate in a research study. I have had the opportunity to ask questions and have had them answered to my satisfaction. I voluntarily agree to participate in this study.

I am not giving up any legal rights by signing this form. I will be given a copy of this form upon request**.**

Printed name of subject	**Signature of subject**
	AM/PM
	Date and time
Printed name of person authorized to consent for subject (when applicable)	**Signature of person authorized to consent for subject (when applicable)**
	AM/PM
Relationship to the subject	**Date and time**

<u>Investigator/Research Staff</u>

I have explained the research to the participant or his/her representative before requesting the signature(s) above. There are no blanks in this document. A copy of this form has been given to the participant or his/her representative.

Printed name of person obtaining consent	**Signature of person obtaining consent**
	AM/PM
	Date and time

List of publications

Publications 2018

Duffy, G., Power, J., Sorby, S. A., & Bowe, B. (2018). Differentiating between spatial ability as a specific rather than general factor of intelligence in performance on simple, non-routine problems in mathematics. Presented at the 72nd Midyear Meeting of the Engineering Design Graphics Division of ASEE, ASEE.

Publications 2017

Duffy, G., Sorby, S. A., Bowe, B., & Nozaki, S. (2017). Evaluation of change in approach to problem solving through developing spatial thinking. Presented at the 45th SEFI Annual Conference, Terceira Island, Azores, Portugal: SEFI.

Duffy, G., Sorby, S. A., Mack, A., & Bowe, B. (2017). Performance by Gender on University Placement Tests in Mathematics and Spatial Skills. Presented at the ASEE Annual Conference 2017, Columbus Ohio: American Society for Engineering Education. Retrieved from https://peer.asee.org/28737.pdf

Publications 2016

Duffy, G., Sorby, S. A., Nozaki, S., & Bowe, B. (2016). Exploring the role of spatial cognition in problem solving. In Frontiers in Education Conference (FIE), 2016 IEEE (pp. 1–4). IEEE. Retrieved from http://ieeexplore.ieee.org/abstract/document/7757593/

Malmi, L., Adawi, T., Curmi, R., de Graaff, E., Duffy, G., Kautz, C., … Williams, B. (2016). How authors did it – a methodological analysis of recent engineering education research papers in the European Journal of Engineering Education. *European Journal of Engineering Education*, 1–19.

Duffy, G., Sorby, S. A., & Bowe, B. (2016). Visualizing Electric Circuits: The Role of Spatial Visualization Skills in Electrical Engineering. Presented at the Proceedings of the 70th Midyear Meeting of the Engineering Design Graphics Division of ASEE, ASEE.

Chance, S. M., Marshall, J., & Duffy, G. (2016). Using Architecture Design Studio Pedagogies to Enhance Engineering Education. International Journal of Engineering Education, 32(1(B)), 364–383.

Publications 2015

G. Duffy, S. Farrell, R. Harding, E. Nevin, A. Mac Raighne, R. Howard, et al. "The effects of spatial skills and spatial skills training on academic performance in STEM education" presented at the Research in Engineering Education Symposium, 2015.

G. Duffy, A. O'Dwyer. "Measurement of first year engineering students cognitive activities using a spatial skills test and an electrical concepts test: implications for curriculum design" presented at the Research in Engineering Education Symposium, 2015.

Publications 2014

G. Duffy and B. Bowe, "An analysis of consistency in, and between, ontology, epistemology and philosophical perspective as contained studies of the first year experience," presented at the Frontiers in Education Conference (FIE), 2014 IEEE, 2014.

Publications 2013

G. Duffy, T. Burke, S. M. Chance, B. Bowe and R. Howard, "Student Experiences of a Project-Based Learning Module" presented at the Proceedings of the 41st SEFI Annual Conference, Leuven, Belgium, 2013

L. Malmi, E. De Graaff, T. Adawi, R. Curmi, G. Duffy, C. Kautz*, et al.*, "Methodological Analysis of SEFI EER papers", presented at the 41st SEFI Annual Conference, 2013

S. M. Chance, G. Duffy, B. Bowe and M. Murphy, "Learning from Learning Groups" presented at the 41st SEFI Annual Conference, 2013.

S. Chance, M. Murphy, G. Duffy, B. Bowe, “Using architecture pedagogy to enhance engineering education” 4th International Research Symposium on Problem-Based Learning (IRSPBL) 2013

Publications 2012

L. Malmi, E. De Graaff, T. Adawi, R. Curmi, G. Duffy, C. Kautz, et al., "Developing a Methodological Taxonomy of EER papers," presented at the 40th SEFI Annual Conference, Thessaloniki, Greece, 2012.

G. Duffy, S. M. Chance, and B. Bowe, "Improving engineering students’ design skills in a project-based learning course by addressing epistemological issues," presented at the 40th SEFI Annual Conference, Thessaloniki, Greece, 2012.

Publications 2011

G. Duffy, "Facilitating intellectual and personal skills development in engineering programmes," presented at the Research in Engineering Education Symposium, Madrid, 2011.

Publications 2010

G. Duffy and B. Bowe, "A strategy for the development of lifelong learning and personal skills throughout an undergraduate engineering programme," presented at the Transforming Engineering Education: Creating Interdisciplinary Skills for Complex Global Environments, Dublin, Ireland, 2010.

G. Duffy and B. Bowe, "A framework to develop lifelong learning and transferable skills in an engineering programme," presented at the 3rd International Symposium for Engineering Education, University College Cork, 2010.